

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 16:59:45 ; Search time 75 Seconds
(without alignments)
365.995 Million cell updates/sec

Title: US-09-887-855-5
Perfect score: 1115
Sequence: 1 ATGRLSGQPVCRGGTQRPC.....EEDAKTKFKESREALNLAY 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1115	100.0	374	21	AAY93948
2	1115	100.0	374	22	AAE03651
3	1115	100.0	374	23	ABG66680
4	1115	100.0	374	23	AB90203
5	1115	100.0	387	22	AA25796
6	1105	99.1	374	21	AA91490
7	1101	98.7	382	20	AAU13367
8	1101	98.7	382	22	AAU29033
9	1101	98.7	382	22	AA80235
10	561.5	50.4	273	21	AA18913

11	561.5	50.4	273	22	AAU12441
12	561.5	50.4	273	22	AA873309
13	561.5	50.4	273	22	AA887609
14	561.5	50.4	273	23	AB985554
15	561.5	50.4	273	23	AB884948
16	504.5	45.2	232	22	AA894192
17	275	24.7	102	22	AB821788
18	275	24.7	102	22	AA229941
19	240	21.5	81	21	AA91643
20	240	21.5	82	21	AA91557
21	240	21.5	115	21	AA91642
22	183	16.4	1456	13	AA824033
23	182	16.3	1184	22	ABG20726
24	178.5	16.0	1455	21	AA823824
25	176.5	15.8	250	22	AAU19630
26	176.5	15.8	250	23	ABP47850
27	176.5	15.8	1479	18	AAW44119
28	176.5	15.8	1479	21	AA843151
29	176.5	15.8	1479	21	AA823823
30	176.5	15.8	1479	23	AB890753
31	175.5	15.7	154	22	AAU19789
32	175.5	15.7	154	23	ABP48009
33	175	15.7	290	22	ABG20724
34	174	15.6	655	23	AA852242
35	174	15.6	741	21	AA843485
36	174	15.6	1257	15	AA845627
37	174	15.6	2409	12	AA812609
38	170.5	15.3	1479	18	AAW44118
39	170.5	15.3	1479	21	AA823822
40	170.5	15.3	1479	23	AB890787
41	165.5	14.8	742	22	AA863347
42	158	14.2	325	23	AAU88024
43	158	14.2	325	23	AAU88026
44	153.5	13.8	637	21	AA832346
45	153.5	13.8	644	21	AAU79186

ALIGNMENTS

RESULT 1
AAY93948
ID AAY93948 standard; Protein; 374 AA.
AC AAY93948;
DT 03-OCT-2000 (first entry)
DE Amino acid sequence of a lectin ss3939 polypeptide.
KW Human; lectin ss3939; chromosome 11; gene therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..21
FT Domain /note= "signal peptide"
FT Domain 22..227
FT Domain /note= "extracellular coding region"
FT Domain 228..248
FT Domain /note= "predicted transmembrane domain"
FT Domain 249..374
FT Domain /note= "predicted cytoplasmic or intracellular domain"

WO2000039296-A1.

06-JUL-2000.

22-DEC-1999; 99WO-US30523.

23-DEC-1998; 98US-0113820.

(IMMV) IMMUNEX CORP.

```

XX PI Anderson DA;
XX DR WPI: 2000-452394/39.
XX DR N-PSDB; AAA57382.
XX
XX PT ss3939 nucleic acids, polypeptides and antibodies, useful for
XX PT identifying human chromosome 11 and diseases associated with it -
XX PS
XX PS Claim 12; Page 8; 73pp; English.
XX
XX CC The present sequence represents a human lectin ss3939 polypeptide. The
XX CC polynucleotide sequence is a source of probes, which may be used
XX CC to identify nucleic acids encoding ss3939 proteins, to identify human
XX CC chromosome number 11, to map genes on human chromosome number 11, to
XX CC identify diseases associated with chromosome 11, as single-stranded
XX CC sense or antisense oligonucleotides to inhibit expression of
XX CC polypeptides encoded by the ss3939 gene, and for gene therapy. The
XX CC ss3939 polypeptides may be useful for developing treatments for
XX CC diseases (none specified) associated with defective or insufficient
XX CC amounts of the polypeptides. The antibodies may be useful for
XX CC detecting the presence of ss3939 polypeptides.
XX
XX SQ Sequence 374 AA;

Query Match 100.0%; Score 1115; DB 21; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.5e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLSSQPCVCGGTGTCPCVYFHDTSRRLNFEAKEACRRDGGQLYSIEDQKL 60
DB 22 ATGRLSSQPCVCGGTGTCPCVYFHDTSRRLNFEAKEACRRDGGQLYSIEDQKL 81
QY 61 IEKFIENLLPSDGFWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPCSGSEV 120
DB 82 IEKFIENLLPSDGFWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPCSGSEV 141
QY 121 CVVMYHOPSPAGTGGPYMFWQNDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 180
DB 142 CVVMYHOPSPAGTGGPYMFWQNDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 201
QY 181 LPEETQEDAKTKFKESREAAALNLAY 206
DB 202 LPEETQEDAKTKFKESREAAALNLAY 227

RESULT 2
AAE03651
ID AAE03651 standard; Protein; 374 AA.
AC AAE03651;
XX
XX DT 06-AUG-2001 (first entry)
XX
XX DE Human extracellular matrix and cell adhesion molecule-15 (XMAD-15).
XX
XX KW Human; extracellular matrix and cell adhesion molecule; XMAD;
XX KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
XX KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
XX KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
XX KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
XX KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
XX KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
XX KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
XX KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
XX KW infection; cell proliferative disorder; actinic keratosis; myeloma;
XX KW arteriosclerosis; neutropenic; anticonvulsant; antithyroid; nephrotropic;
XX KW neuroprotective; dermatological.
XX
XX OS Homo sapiens.
XX
XX FT Key Location/Qualifiers
XX FT Peptide 1..24

```

```

FT Protein /label= Signal_peptide
FT 25..374 /note= "Mature human extracellular matrix and cell
FT adhesion molecule (XMAD)".
FT 46..63 /note= "C-type lectin domain"
FT 163..176 /note= "C-type lectin domain"
FT 224..247 /note= "Transmembrane motif"
FT 328..348 /note= "Transmembrane motif"
FT
FT WO200142285-A2.
FT 14-JUN-2001.
XX
XX 05-DEC-2000; 2000MO-US32990.
XX
XX 10-DEC-1999; 99US-0172852.
XX
XX 16-DEC-1999; 99US-0172354.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
XX Baughn MR, Lu DAM, Shah P, Au-Young J;
XX
XX WPI: 2001-381632/40.
XX N-PSDB; AAD08059.
XX
XX New human extracellular matrix and cell adhesion molecules and
XX polynucleotide sequences encoding them, useful for diagnosis,
XX prevention, treatment of genetic, autoimmune and cell proliferative
XX disorders
XX
XX Claim 1; Page 108-109; 135pp; English.
XX
XX The present sequence is a human extracellular matrix and cell
XX adhesion molecule (XMAD). The XMAD is used for screening a compound for
XX effectiveness as an agonist or antagonist of XMAD. The identified agonist
XX or antagonist are used for treating a disease or condition associated
XX with decreased or increased expression of functional XMAD. The
XX polynucleotides encoding XMAD are useful in somatic or germine gene
XX therapy to correct a genetic deficiency, to express a conditionally
XX lethal gene product and to express a protein which affords protection
XX against intracellular parasites and also for diagnosis of disorders
XX associated with expression of XMAD. They are also used for generating
XX hybridisation probes useful in mapping the naturally occurring genomic
XX sequences and to create knock in humanised animals (pigs) or transgenic
XX animals (mice or rats) to model human diseases. Oligonucleotide or longer
XX fragments derived from the polynucleotide sequences may be used as
XX elements on a microarray. Antibodies which specifically bind XMAD may be
XX used for the diagnosis of disorders associated with the expression of
XX XMAD, or in assays to monitor patients being treated with XMAD. Diseases
XX diagnosed, prevented or treated include genetic disorders such as
XX adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
XX disease, myotonic dystrophy, sickle cell anaemia, thalassaemia,
XX autoimmune/inflammatory disorders such as acquired immune deficiency
XX syndrome (AIDS), Addison's disease, diabetes mellitus, atopic dermatitis,
XX atherosclerosis, Crohn's disease, allergies, anaemia, asthma,
XX glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
XX bacterial, fungal, parasitic, rheumatoid arthritis, ulcerative colitis,
XX cell proliferative disorders such as actinic keratosis, arteriosclerosis
XX and cancer including breast, bladder, bone marrow, brain and uterus
XX cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.
XX
XX SQ Sequence 374 AA;

Query Match 100.0%; Score 1115; DB 22; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.5e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGRLLSGQPVCRGGTQRCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSISEDEQKL 60
|||||
Db 22 ATGRLLSGQPVCRGGTQRCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSISEDEQKL 81
QY 61 IEKFTIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMVYDPSGSEV 120
Db 82 IEKFTIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMVYDPSGSEV 141
QY 121 CVVMYHQPAPAGIGGYPYFQWDDRCNKNMFICKYSDEKPAVPSREAEGETELTTPV 180
Db 142 CVVMYHQPAPAGIGGYPYFQWDDRCNKNMFICKYSDEKPAVPSREAEGETELTTPV 201
QY 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227
RESULT 3
ABG66680 standard; Protein: 374 AA.
ID ABG66680
XX AC ABG66680;
DT 30-AUG-2002 (first entry)
XX DE Human novel polypeptide #15.
XX KW Human; inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; hematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW fungal infection.
XX OS Homo sapiens.
XX PN WO200244340-A2.
XX PD 06-JUN-2002.
XX PF 30-NOV-2001; 2001WO-US47004.
XX PR 30-NOV-2000; 2000US-0028952.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI Yamazaki V, Ujwal ML, Drmanac RT;
XX N-PSDB; ABK94904.
XX WPI; 2002-508509/54.
XX DR Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing -
XX Claim 10; Page 579-580; 672pp; English.
CC The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia,

CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABG66666-ABG666758 represent human
XX novel polypeptides of the invention.
XX Sequence 374 AA;
Query Match 100.0%; Score 1115; DB 23; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.5e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGRLLSGQPVCRGGTQRCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSISEDEQKL 60
Db 22 ATGRLLSGQPVCRGGTQRCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSISEDEQKL 81
QY 61 IEKFTIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMVYDPSGSEV 120
Db 82 IEKFTIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMVYDPSGSEV 141
QY 121 CVVMYHQPAPAGIGGYPYFQWDDRCNKNMFICKYSDEKPAVPSREAEGETELTTPV 180
Db 142 CVVMYHQPAPAGIGGYPYFQWDDRCNKNMFICKYSDEKPAVPSREAEGETELTTPV 201
QY 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227
RESULT 4
ABG90203 standard; Protein: 374 AA.
ID ABB90203
XX AC ABB90203;
DT 24-MAY-2002 (first entry)
XX DE Human polypeptide SEQ ID NO 2579.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX OS Homo sapiens.
XX PN WO200190304-A2.
XX PD 29-NOV-2001.
XX PF 18-MAY-2001; 2001WO-US16450.
XX PR 19-MAY-2000; 2000US-205515P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Birse CE, Rosen CA;
XX WPI; 2002-122018/16.
XX DR N-PSDB; ABL90612.
XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX disorders -

PS Claim 11; SEQ ID NO 2579; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB8040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

XX Sequence 374 AA;

Query Match 100.0%; Score 1115; DB 23; Length 374;

Best Local Similarity 100.0%; Pred. No. 3.5e-105;

Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLSGQPCVCGTQPCYKVIFYHDTSRRLNFEAKEACRRDGGOLVSEDEQKL 60

DB 22 ATGRLSGQPCVCGTQPCYKVIFYHDTSRRLNFEAKEACRRDGGOLVSEDEQKL 81

QY 61 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLIYAWTDGSIQFRNMYVDEPSCGSEV 120

DB 82 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLIYAWTDGSIQFRNMYVDEPSCGSEV 141

QY 121 CVVWHOPSPAGTGGPYMFOWNDRCKMKNFKICKYSDKPAVPSRAEGEETELTPV 180

DB 142 CVVWHOPSPAGTGGPYMFOWNDRCKMKNFKICKYSDKPAVPSRAEGEETELTPV 201

QY 181 LPEETQEDAKKTFKESREAAALNAY 206

DB 202 LPEETQEDAKKTFKESREAAALNAY 227

RESULT 5

AAM25796

ID AAM25796 standard; Protein: 387 AA.

XX

XX AAM25796;

DT 16-OCT-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:1311.

XX

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema; dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder.

OS Homo sapiens.

XX

PN WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000MO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-457603/49.

DR N-PSDB; AAH99737.

XX Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection.

XX

PS Claim 20; Page 272; 1217pp; English.

XX

CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmune, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopaenia, wounds, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

XX

SQ Sequence 387 AA;

Query Match 100.0%; Score 1115; DB 22; Length 387;

Best Local Similarity 100.0%; Pred. No. 3.6e-105;

Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLSGQPCVCGTQPCYKVIFYHDTSRRLNFEAKEACRRDGGOLVSEDEQKL 60

DB 35 ATGRLSGQPCVCGTQPCYKVIFYHDTSRRLNFEAKEACRRDGGOLVSEDEQKL 94

QY 61 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLIYAWTDGSIQFRNMYVDEPSCGSEV 120

DB 95 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLIYAWTDGSIQFRNMYVDEPSCGSEV 154

QY 121 CVVWHOPSPAGTGGPYMFOWNDRCKMKNFKICKYSDKPAVPSRAEGEETELTPV 180

DB 155 CVVWHOPSPAGTGGPYMFOWNDRCKMKNFKICKYSDKPAVPSRAEGEETELTPV 214

QY 181 LPEETQEDAKKTFKESREAAALNAY 206

DB 215 LPEETQEDAKKTFKESREAAALNAY 240

RESULT 6

AA91490

ID AA91490 standard; Protein: 374 AA.

XX

AC AA91490;

XX

DT 29-JUN-2000 (first entry)

XX DE Human secreted protein sequence encoded by gene 40 SEQ ID NO:163.

KW Human: secreted protein; diagnosis: cytostatic; immunosuppressive;

KW antiHIV; antiinflammatory; neutropic; neuroprotective; antiallergic;

KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;

KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;

KW immune disease; inflammation; blood disorder; tumour.

XX OS Homo sapiens.

XX PN WO200006698-A1.

XX PD 10-FEB-2000.

XX PF 29-JUL-1999; 99WO-US17130.

XX PR 30-JUL-1998; 98US-0094657.

XX PR 05-AUG-1998; 98US-0095486.

XX PR 06-AUG-1998; 98US-0095454.

XX PR 06-AUG-1998; 98US-0095455.

XX PR 12-AUG-1998; 98US-0096319.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;

PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;

PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;

XX DR WPI; 2000-195282/17.

XX DR N-P5DB; AAA26385.

XX PT New isolated human genes and the secreted polypeptides they encode,

PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX PS Claim 11; Page 483-484; 634pp; English.

XX CC The polynucleotide sequences given in AAA26346 to AAA26458 encode the

CC human secreted proteins given in AAY91451 to AAY91691. The human secreted

CC proteins can have activities based on the tissues and cells they are

CC expressed in. Examples of the activities are: cytostatic;

CC immunosuppressive; antiHIV; antiinflammatory; neutropic; neuroprotective;

CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;

CC antiasthma; antipsoriatic; and cardiant. The polynucleotides and their

CC corresponding secreted proteins are useful for preventing, treating or

CC ameliorating medical conditions, e.g. by protein or gene therapy. Also

CC pathological conditions can be diagnosed by determining the amount of the

CC proteins in a sample or by determining the presence of mutations in the

CC polynucleotides. Specific uses are described for each of the

CC polynucleotides, based on which tissues they are most highly expressed

CC in, and include developing products for the diagnosis or treatment of

CC cancer, tumours, neurodegenerative disorders, developmental abnormalities

CC and foetal deficiencies, blood disorders, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, inflammation,

CC allergies, Alzheimer's and behavioural disorders, schizophrenia,

CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,

CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,

CC cardiovascular disorders, reproductive disorders, gastrointestinal

CC disorders, respiratory disorders and metabolic disorders. The proteins

CC or polynucleotides can also be used as food additives or preservatives.

CC The proteins are also useful for identifying their binding partners.

CC AAA26337 to AAA26345 and AAY91450 are sequences used in the

CC exemplification of the present invention.

XX SQ Sequence 374 AA;

Query Match 99.18; Score 1105; DB 21; Length 374;

Best Local Similarity 99.08; Pred. No. 3.6e-104;

Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGRLSGPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSIEDEQKL 60

|||||

Db 22 ATGRLSGPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSIEDEQKL 81

QY 61 IEKFIENLLPSDGFDFWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYDEPSCSEV 120

Db 82 IEXFIENLLPSDGFDFWGLRRREKQSNSTXCQDLYAWTDGSIQFRNYYDEPSCSEV 141

QY 121 CVVMYHQPSAPAGIGGYPYFQWNDRCNMKNFNICKYSDEKPAVPSREAGEETEELTPV 180

Db 142 CVVMYHQPSAPAGIGGYPYFQWNDRCNMKNFNICKYSDEKPAVPSREAGEETEELTPV 201

QY 181 LPEETOEDAKTKFKESREAAALNLAY 206

Db 202 LPEETOEDAKTKFKESREAAALNLAY 227

RESULT 7

AAY13367

ID AAY13367 standard; Protein; 382 AA.

XX AC AAY13367;

XX DT 25-JUN-1999 (first entry)

XX DE Amino acid sequence of protein PRO234.

XX KW Secreted protein; transmembrane protein; human; enterocolitis;

KW Zollinger-Ellison syndrome; gastrointestinal ulceration;

KW congenital microvillus atrophy; skin disease; cell growth;

KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;

KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;

KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;

KW anti-thrombotic; wound healing; tissue repair.

XX OS Homo sapiens.

XX PN WO9914328-A2.

XX PD 25-MAR-1999.

XX PF 16-SEP-1998; 98WO-US19330.

XX PR 25-NOV-1997; 97US-0066840.

PR 17-SEP-1997; 97US-0059113.

PR 17-SEP-1997; 97US-0059115.

PR 17-SEP-1997; 97US-0059117.

PR 17-SEP-1997; 97US-0059119.

PR 17-SEP-1997; 97US-0059121.

PR 17-SEP-1997; 97US-0059122.

PR 18-SEP-1997; 97US-0059184.

PR 18-SEP-1997; 97US-0059263.

PR 15-OCT-1997; 97US-0059266.

PR 15-OCT-1997; 97US-0062125.

PR 17-OCT-1997; 97US-0062285.

PR 21-OCT-1997; 97US-0062287.

PR 21-OCT-1997; 97US-0063486.

PR 24-OCT-1997; 97US-0062814.

PR 24-OCT-1997; 97US-0062816.

PR 24-OCT-1997; 97US-0063045.

PR 24-OCT-1997; 97US-0063120.

PR 24-OCT-1997; 97US-0063121.

PR 24-OCT-1997; 97US-0063127.

PR 24-OCT-1997; 97US-0063128.

PR 27-OCT-1997; 97US-0063329.

PR 27-OCT-1997; 97US-0063327.

PR 28-OCT-1997; 97US-0063541.

PR 28-OCT-1997; 97US-0063542.

PR 28-OCT-1997; 97US-0063544.

PR 28-OCT-1997; 97US-0063549.

PR 28-OCT-1997; 97US-0063550.

PR 28-OCT-1997; 97US-0063564.

PR 29-OCT-1997; 97US-0063435.

PR 29-OCT-1997; 97US-0063704.

PR 29-OCT-1997; 97US-0063732.

PR 29-OCT-1997; 97US-0063738.
PR 29-OCT-1997; 97US-0063734.
PR 29-OCT-1997; 97US-0064215.
PR 29-OCT-1997; 97US-0063735.
PR 31-OCT-1997; 97US-0063870.
PR 31-OCT-1997; 97US-0064103.
PR 03-NOV-1997; 97US-0064248.
PR 07-NOV-1997; 97US-0064809.
PR 12-NOV-1997; 97US-0065186.
PR 17-NOV-1997; 97US-0065846.
PR 18-NOV-1997; 97US-0065693.
PR 21-NOV-1997; 97US-0066120.
PR 21-NOV-1997; 97US-0066364.
PR 24-NOV-1997; 97US-0066772.
PR 24-NOV-1997; 97US-0066466.
PR 24-NOV-1997; 97US-0066770.
PR 24-NOV-1997; 97US-0066511.
PR 24-NOV-1997; 97US-0066453.
XX (GETH) GENENTECH INC.
PA
XX
XX
PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
XX N-PSDB: AAX52238.
DR WPI: 1999-229533/19.
DR N-PSDB: AAX52238.
XX
XX
PT New isolated human genes and polypeptides used in, e.g. treatment of
PT gastrointestinal ulceration
XX
XX
XX Claim 12: Fig 50; 320pp; English.
XX
XX
CC AAY13344-403 represent secreted and transmembrane human proteins.
CC The cDNA sequences are obtained from cDNA libraries, prepared from
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC The encoded polypeptides have specific uses based on their homology to
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC associated with the preservation and maintenance of gastrointestinal
CC mucosa and the repair of acute and chronic mucosal lesions
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC ulceration and congenital microvillus atrophy), skin diseases associated
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC potent effects on cell growth and development, diseases related to
CC growth or survival of nerve cells including Parkinson's disease,
CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC anti-thrombotic agent; PRO287 polypeptides and portions may have
CC therapeutic applications in wound healing and tissue repair; PRO317 can
CC be used for treating problems of the kidney, uterus, endometrium, blood
CC vessels, or related tissue, e.g. in the heart of genital tract.
XX
XX
SQ Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 20; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.5e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSI 52
|||||
22 ATGRLLSASDLRLGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSI 81
|||||
QY 53 ESEDEQKLIKFTENLLPSDGFGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVD 112
|||||
82 ESEDEQKLIKFTENLLPSDGFGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVD 141
|||||
QY 113 EPSGSGEVCVVMYHQPSAPAGIGPYMFQWDDRCNMKNNFICKYSDEKPAVPSRAEGE 172
|||||
142 EPSGSGEVCVVMYHQPSAPAGIGPYMFQWDDRCNMKNNFICKYSDEKPAVPSRAEGE 201
|||||
QY 173 ETELTPVLPEETQEDAKKTFKESREALNLAY 206
|||||

DB 202 ETELTPVLPEETQEDAKKTFKESREALNLAY 235
RESULT 8
AAU29033
ID AAU29033 standard; Protein; 382 AA.
XX
XX AAU29033;
XX
XX 18-DEC-2001 (first entry)
XX Human PRO polypeptide sequence #10.
XX
XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
XX dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
XX blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
XX adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200168848-A2.
XX
XX 20-SEP-2001.
XX
XX 28-FEB-2001; 2001WO-US06520.
XX
XX 01-MAR-2000; 2000WO-US05601.
XX 02-MAR-2000; 2000WO-US05841.
XX 03-MAR-2000; 2000US-187202P.
XX 06-MAR-2000; 2000US-186968P.
XX 14-MAR-2000; 2000US-189320P.
XX 15-MAR-2000; 2000US-189328P.
XX 15-MAR-2000; 2000WO-US06884.
XX 21-MAR-2000; 2000US-190828P.
XX 21-MAR-2000; 2000US-191007P.
XX 21-MAR-2000; 2000US-191048P.
XX 21-MAR-2000; 2000US-191314P.
XX 28-MAR-2000; 2000US-192655P.
XX 29-MAR-2000; 2000US-193032P.
XX 29-MAR-2000; 2000US-193053P.
XX 30-MAR-2000; 2000WO-US08439.
XX 04-APR-2000; 2000US-194449P.
XX 11-APR-2000; 2000US-195975P.
XX 11-APR-2000; 2000US-196000P.
XX 11-APR-2000; 2000US-196187P.
XX 11-APR-2000; 2000US-196690P.
XX 11-APR-2000; 2000US-196820P.
XX 18-APR-2000; 2000US-198121P.
XX 18-APR-2000; 2000US-198585P.
XX 25-APR-2000; 2000US-199397P.
XX 25-APR-2000; 2000US-199550P.
XX 25-APR-2000; 2000US-199654P.
XX 03-MAY-2000; 2000US-201516P.
XX 17-MAY-2000; 2000WO-US13705.
XX 22-MAY-2000; 2000WO-US14042.
XX 30-MAY-2000; 2000WO-US14941.
XX 02-JUN-2000; 2000WO-US15264.
XX 05-JUN-2000; 2000US-209832P.
XX 28-JUL-2000; 2000WO-US20710.
XX 22-AUG-2000; 2000US-064848.
XX 24-AUG-2000; 2000WO-US23328.
XX 08-NOV-2000; 2000WO-US30952.
XX 01-DEC-2000; 2000WO-US32678.
XX 20-DEC-2000; 2000WO-US34956.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
XX Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI: 2001-602746/68.
XX N-PSDB: AAS45934.
DR
DR

XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
PS Claim 11; Fig 20; 77app; English.
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
CC The PRO polypeptides and their associated nucleic acids can be used to
CC detect the presence of a tumour in a mammal by comparing the level of
CC expression of a PRO polypeptide in a test sample of cells from the animal
CC and a control sample of normal cells, whereby a higher level of
CC expression in the test sample indicates the presence of a tumour in the
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC and rabbits but are preferably human. The polypeptides can be used to
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC when contacted with it. A specific polypeptide can be used to stimulate
CC the proliferation or differentiation of chondrocyte cells. The PRO
CC proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
XX Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 22; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.5e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 52
DB 22 ATGRLLSADLDLRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 81
QY 53 ESEDEQKLEKFTENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNNYVD 112
DB 82 ESEDEQKLEKFTENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNNYVD 141
QY 113 EPSCGSEVCVMYHQPAPAGIGGYPYMFQWDDRCNKNFNICKYSDKPAVPSREAAGE 172
DB 142 EPSCGSEVCVMYHQPAPAGIGGYPYMFQWDDRCNKNFNICKYSDKPAVPSREAAGE 201
QY 173 ETELTTPVLPEETQEDAKKTFKESREAAALNLAY 206
DB 202 ETELTTPVLPEETQEDAKKTFKESREAAALNLAY 235

RESULT 9
AAB80235
ID AAB80235 standard; Protein; 382 AA.
XX AAB80235;
AC AAB80235;
DT 24-APR-2001 (first entry)
XX Human PRO234 protein.
DE Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
XX antiparkinsonian nootropic; neuroprotective; vulnerrary; cardiant;
KW antilangogenic; vasotropic; antiasthmatic; antirheumatic; cancer;
KW antilarthritic; antiinfertility; antidiabetic; antiviral; diabetes;
KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
XX ischaemia; inflammation.
XX Homo sapiens.
OS WO200104311-A1.
PN 18-JAN-2001.
XX 22-FEB-2000; 2000WO-US04414.
XX 07-JUL-1999; 99US-0143048.
PR

PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 05-JAN-2000; 99WO-US00219.
XX (GETH) GENENTECH INC.

PA Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
PI Flivaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;
PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI Williams PM, Wood WI;
XX WPI; 2001-081051/09.
DR N-PSDB; AAF72396.

Sixty one nucleic acids encoding PRO polypeptides which are useful in
the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
PT squamous cell carcinoma) and neurodegenerative diseases (e.g.
PT Alzheimer's disease) -

Claim 1; Fig 50; 393pp; English.

XX The present sequence is one of sixty one novel secreted and
XX transmembrane PRO polypeptides. The PRO polypeptides are
XX useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
CC squamous cell carcinoma), gastrointestinal disorders (e.g.
CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.
CC endometrial bleeding angiogenesis, ischaemia such as coronary
CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
CC diabetes and retinal disorders such as retinitis pigmentosum.
CC The PRO nucleic acids have applications in molecular biology, including
XX use as hybridization probes, and in chromosome and gene mapping.

Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 22; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.5e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 52
DB 22 ATGRLLSADLDLRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 81
QY 53 ESEDEQKLEKFTENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNNYVD 112
DB 82 ESEDEQKLEKFTENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNNYVD 141
QY 113 EPSCGSEVCVMYHQPAPAGIGGYPYMFQWDDRCNKNFNICKYSDKPAVPSREAAGE 172
DB 142 EPSCGSEVCVMYHQPAPAGIGGYPYMFQWDDRCNKNFNICKYSDKPAVPSREAAGE 201
QY 173 ETELTTPVLPEETQEDAKKTFKESREAAALNLAY 206
DB 202 ETELTTPVLPEETQEDAKKTFKESREAAALNLAY 235

RESULT 10
AAB18913
ID AAB18913 standard; Protein; 273 AA.
XX AAB18913;
AC AAB18913;

```

XX DT 08-FEB-2001 (first entry)
XX DE A novel polypeptide designated PRO1890.
XX KW Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
KW PRO1890; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
KW PRO4352; PRO4380; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
KW PRO4424; PRO4422; PRO4430; PRO4499; tumour; obesity; diabetes;
KW Insulinemia; kidney disorder; Bergers disease; nephropathy;
KW Schonelein-Henoch purpura; celliac disease; dermatitis herpeticiformis;
KW Crohns disease.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FH Peptide 1..21
FT Modified-site /note= "signal peptide"
FT Modified-site 27..32
FT Modified-site /note= "N-myristoylation site"
FT Modified-site 66..71
FT Misc-difference /note= "N-myristoylation site"
FT /note= "Glu encoded by CAA"
FT Modified-site 91..96
FT Modified-site /note= "N-myristoylation site"
FT Modified-site 93..98
FT Modified-site /note= "N-myristoylation site"
FT Modified-site 102..107
FT Modified-site /note= "N-myristoylation site"
FT Modified-site 109..114
FT Modified-site /note= "N-myristoylation site"
FT Modified-site 140..145
FT Modified-site /note= "N-myristoylation site"
FT Modified-site 212..217
FT Domain /note= "N-myristoylation site"
FT /note= "transmembrane domain"
FT Modified-site 266..269
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
XX WO200056889-A2.
XX PD 28-SEP-2000.
XX PF 01-MAR-2000; 2000WO-US05601.
XX 23-MAR-1999; 99US-0125774.
XX 23-MAR-1999; 99US-0125778.
XX 24-MAR-1999; 99US-0125826.
XX 31-MAR-1999; 99US-0127035.
XX 03-APR-1999; 99US-0127706.
XX 21-APR-1999; 99US-0130359.
XX 27-APR-1999; 99US-0131270.
XX 27-APR-1999; 99US-0131272.
XX 27-APR-1999; 99US-0131291.
XX 04-MAY-1999; 99US-0132371.
XX 04-MAY-1999; 99US-0132379.
XX 04-MAY-1999; 99US-0132383.
XX 25-MAY-1999; 99US-0135750.
XX 08-JUN-1999; 99US-0138166.
XX 20-JUL-1999; 99US-0144791.
XX 03-AUG-1999; 99US-0146970.
XX 09-DEC-1999; 99US-0170262.
XX (GETH ) GENENTECH INC.
XX Desnoyers L, Eaton DL, Goddard A, Godowski RJ, Gurney AL, Pan J;
PI Stewart TA, Watanabe CK, Wood WI, Zhang Z;
XX WPI; 2000-628263/60.
XX DR N-PSDB; AAA96340.

```

```

XX Novel secreted and transmembrane polypeptides useful for diagnosing
PT tumour in a mammal, for identifying agonists and antagonists of the
PT polypeptide and for therapeutic use
XX Claim 12; Fig 10; 222pp; English.
XX The present sequence represents a secreted or transmembrane polypeptide.
XX The specification describes polypeptides designated PRO1484, PRO4334,
CC PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
CC PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
CC PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
CC useful for diagnosing tumour in a mammal. The polypeptides, their
CC agonists and antagonists are useful treating a condition associated with
CC expression or activity of the polypeptide. Conditions treated include
CC obesity, diabetes or hyper- or hypo-insulinemia. The polypeptides are
CC capable of inducing proliferation of mammalian kidney mesangial cells
CC and are therefore useful for treating kidney disorders associated with
CC decreased mesangial cell function such as Bergers disease or other
CC nephropathies associated with Schonelein-Henoch purpura, celliac disease,
CC dermatitis herpeticiformis or Crohns disease. The nucleic acids may be used
CC to generate transgenic animals for use in development and screening of
CC therapeutically useful reagents and also for chromosome identification
CC and tissue typing.
XX Sequence 273 AA;
XX Query Match 50.4%; Score 561.5; DB 21; Length 273;
XX Best Local Similarity 60.1%; Pred. No. 7e-49;
XX Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;
QY 4 RLLSGQPVCRGRTQPCYKVIYFHDTRRLNFEPAKEACRRDGGOLVSISEDEOKLIEK 63
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 23 RVSGQKVCFAFDRKPCYKMAFYHELSRVSFQEARLACESEGGVLLSLENEAEOKLIES 82
QY 64 FIENLLP-----SDGDFWIGLRRREKQSNSTACODLYAWTDGSGISQFRNHYVDPSGGS 118
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 83 MLQNTLTPGTGTSIDGDFWIGLRNGDGT--SGACPDLYQWSDGSGNSQYRNWYTDPSGGS 141
QY 119 EVCVYVYHQPAPAGIGPYFMFQWDDRCNNKNNFICKYSDE-KPAVP 165
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 142 EKCVCVYVYHQPANPLGLGPGYLYQWDDRCNNKNNHYICKYEPEINPTAP 189
RESULT 11
AAU12441
ID AAU12441 standard; Protein; 273 AA.
XX AAU12441;
AC AAU12441;
XX 24-OCT-2001 (first entry)
XX Human PRO1890 polypeptide sequence.
XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW adipocyte; A-peptide; factor VIIA; gene therapy.
XX Homo sapiens.
XX WO20010466-A2.
XX 07-JUN-2001.
XX 01-DEC-2000; 2000WO-US32678.
XX 01-DEC-1999; 99WO-US28301.
XX 01-DEC-1999; 99WO-US28634.
XX 02-DEC-1999; 99WO-US28551.
XX 02-DEC-1999; 99WO-US28564.
XX 02-DEC-1999; 99WO-US28565.
XX 09-DEC-1999; 99US-0170262.

```


Best Local Similarity 60.1%; Pred. No. 7e-49; Mismatches 25; Indels 7; Gaps 3;
Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;

QY 4 RLLSQPVCRGQTQPCYKVIYFHDTSRLNFEAEAKERRDGGQLGVSTESDEQKLIK 63
Db 23 RYVSGKVCFAFKPCYKMAFYHELSSRVSFQEARLACESEGGVLLSLENAEQKLIES 82

QY 64 FIENLLP-----SDGDFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGS 118
Db 83 MLQNIITKPGTIGSDGDFWIGLWRNGDGT-SGACPDLYQWSDGNSQYRNMYTDEPSCGS 141

QY 119 EYCVVYHQPAPAGIGGYPYFQWDDRCNMKNFICKYSDE-KPAVP 165
Db 142 EKCVMYHQPANTPGGLGYPYQWDDRCNMKNHNTICKYEPEINPTAP 189

RESULT 13
AAB87609
ID AAB87609 standard; Protein; 273 AA.
XX AAB87609;
XX DT 15-MAY-2001 (first entry)
XX AC AAB87609;
XX DE Human PRO1890.
XX KW Human; PRO protein; mapping.
XX OS Homo sapiens.
XX PN WO2001163118-A2.
XX PD 08-MAR-2001.
XX PF 24-AUG-2000; 2000WO-US23328.
XX PR 01-SEP-1999; 99WO-US20111.
XX PR 15-SEP-1999; 99WO-US21090.
XX PR 07-DEC-1999; 99US-0169495.
XX PR 09-DEC-1999; 99US-0170262.
XX PR 11-JAN-2000; 2000US-0175481.
XX PR 18-FEB-2000; 2000WO-US04341.
XX PR 18-FEB-2000; 2000WO-US04342.
XX PR 22-FEB-2000; 2000WO-US04414.
XX PR 01-MAR-2000; 2000WO-US05601.
XX PR 03-MAR-2000; 2000US-0187202.
XX PR 25-APR-2000; 2000US-0199397.
XX PR 22-MAY-2000; 2000WO-US14042.
XX PR 05-JUN-2000; 2000US-0209832.
XX PA (GETH) GENENTECH INC.
XX PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
XX PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX DR WPI; 2001-183260/18.
XX DR N-PSDB; AAF92141.
XX PT Eighty four nucleic acids encoding PRO polypeptides, useful in
XX PT molecular biology, including use as hybridization probes, and in
XX PT chromosome and gene mapping.
XX PS Claim 12; Fig 168; 278pp; English.
XX CC The present sequence is a human PRO polypeptide (secreted and
XX CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
XX CC anti-PRO antibodies are useful for preparation of a medicament useful in
XX CC the treatment of a condition which is responsive to the PRO protein,
XX CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
XX CC employed as molecular weight markers for protein electrophoresis. The PRO
XX CC coding sequence has applications in molecular biology, including use as
XX CC hybridisation probes, and in chromosome and gene mapping.

SQ Sequence 273 AA;
Query Match 50.4%; Score 561.5; DB 22; Length 273;
Best Local Similarity 60.1%; Pred. No. 7e-49;
Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;

QY 4 RLLSQPVCRGQTQPCYKVIYFHDTSRLNFEAEAKERRDGGQLGVSTESDEQKLIK 63
Db 23 RYVSGKVCFAFKPCYKMAFYHELSSRVSFQEARLACESEGGVLLSLENAEQKLIES 82

QY 64 FIENLLP-----SDGDFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGS 118
Db 83 MLQNIITKPGTIGSDGDFWIGLWRNGDGT-SGACPDLYQWSDGNSQYRNMYTDEPSCGS 141

QY 119 EYCVVYHQPAPAGIGGYPYFQWDDRCNMKNFICKYSDE-KPAVP 165
Db 142 EKCVMYHQPANTPGGLGYPYQWDDRCNMKNHNTICKYEPEINPTAP 189

RESULT 14
AAB95554
ID AAB95554 standard; Protein; 273 AA.
XX AAB95554;
XX AC AAB95554;
XX DT 19-JUL-2002 (first entry)
XX DE Human angiogenesis related protein PRO1890 SEQ ID NO: 264.
XX KW Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
XX KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
XX KW cardiant; cytostatic; antiangiogenic; hypotensive; vulnerary;
XX KW antiarteriosclerotic.
XX OS Homo sapiens.
XX PN WO200208284-A2.
XX PD 31-JAN-2002.
XX PF 09-JUL-2001; 2001WO-US21735.
XX PR 20-JUL-2000; 2000US-219556P.
XX PR 25-JUL-2000; 2000US-220624P.
XX PR 25-JUL-2000; 2000US-220664P.
XX PR 28-JUL-2000; 2000WO-US20710.
XX PR 02-AUG-2000; 2000US-222695P.
XX PR 17-AUG-2000; 2000US-0643657.
XX PR 23-AUG-2000; 2000WO-US23328.
XX PR 24-AUG-2000; 2000WO-US23328.
XX PR 07-SEP-2000; 2000US-230978P.
XX PR 15-SEP-2000; 2000US-000000P.
XX PR 18-SEP-2000; 2000US-0664610.
XX PR 18-SEP-2000; 2000US-0665350.
XX PR 24-OCT-2000; 2000US-242922P.
XX PR 08-NOV-2000; 2000US-0709238.
XX PR 08-NOV-2000; 2000WO-US30952.
XX PR 10-NOV-2000; 2000WO-US30873.
XX PR 01-DEC-2000; 2000WO-US32678.
XX PR 20-DEC-2000; 2000US-0747259.
XX PR 20-DEC-2000; 2000WO-US34956.
XX PR 22-JAN-2001; 2001US-0767609.
XX PR 28-FEB-2001; 2001US-0796498.
XX PR 28-FEB-2001; 2001WO-US06520.
XX PR 01-MAR-2001; 2001WO-US06666.
XX PR 09-MAR-2001; 2001US-0802706.
XX PR 14-MAR-2001; 2001US-0808689.
XX PR 22-MAR-2001; 2001US-0816744.
XX PR 05-APR-2001; 2001US-0828366.
XX PR 10-MAY-2001; 2001US-0854208.
XX PR 10-MAY-2001; 2001US-0854280.
XX PR 25-MAY-2001; 2001US-0866028.
XX PR 25-MAY-2001; 2001US-0866034.

```

PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001US-0870574.
PR 01-JUN-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 28-JUN-2001; 2001WO-US00000.
XX (GETH ) GENENTECH INC.
PA (BAKE/) BAKER K P.
PA (FERR/) FERRARA N.
PA (GERB/) GERBER H.
PA (GERR/) GERRITSEN M E.
PA (GODD/) GODDARD A.
PA (GODO/) GODOWSKI P J.
PA (GURN/) GURNEY A L.
PA (HILL/) HILLAN K J.
PA (MARS/) MARSTERS S A.
PA (PANJ/) PAN J.
PA (PAON/) PAONI N F.
PA (STEP/) STEPHAN J F.
PA (WATA/) WATANABE C K.
PA (WILL/) WILLIAMS P M.
PA (WOOD/) WOOD W I.
XX
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
DR WPI; 2002-171999/22.
DR N-PSDB; ABL95692.
XX
XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
XX useful in diagnosis and treatment of cardiovascular (e.g. myocardial
XX infarction), endothelial or angiogenic disorders in a mammal .
XX
XX Claim 11; Fig 264; 567pp; English.
XX
XX The present invention provides the protein and coding sequences of human
XX PRO proteins. These are useful for treating or diagnosing a
XX cardiovascular, endothelial or angiogenic disorder, including cardiac
XX hypertrophy, trauma, cancer, age-related macular degeneration,
XX atherosclerosis, hypertension, arterial stenosis, rheumatoid arthritis,
XX angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
XX angiogenesis (such as breast carcinoma and liver carcinoma) and wound
XX healing. The present sequence is a PRO protein of the invention.
XX
XX SQ Sequence 273 AA;
XX
XX Query Match 50.4%; Score 561.5; DB 23; Length 273;
XX Best Local Similarity 60.1%; Pred. No. 7e-49;
XX Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;
XX
QY 4 RLLSGQPVCRGQTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSIKIEDEKLIK 63
Db :||| | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
23 RVVSGQKVCFADEKPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSIKIEDEKLIK 82
QY 64 FIENLPL-----SDGDFWIGLRREKQSNSTACODLYAMTDGSIQFRNMYVDEPSCGS 118
Db 83 MLQNLTKPGTGISDGFWIGLRNGDGT-SGACPDLYQWSDGNSQYRNWYTDPSGCS 141
QY 119 EVCVVMYHQPAPAGIGGPFYQWQNDRCNKNFNICKYSDE-KPVP 165
Db :||| | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
142 EKCVMYHQPANPLGLGPLYIQWNDRCNKNFNICKYEPINPTAP 189
XX
RESULT 15
ID ABB84948
XX ABB84948 standard; Protein; 273 AA.
XX
XX ABB84948;
XX
DT 16-MAY-2002 (first entry)
XX

```

```

DE Human PRO1890 protein sequence SEQ ID NO:264.
XX
KW Human; angiogenesis; cardiant; cytostatic; antiangiogenic; hypotensive;
KW vulnary; antiarteriosclerotic; PRO agonist; PRO antagonist; cancer;
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW age-related macular degeneration; arterial stenosis; angina;
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
XX wound healing; chromosome mapping; gene mapping.
XX
OS Homo sapiens.
XX
PN WO200200690-A2.
XX
PD 03-JAN-2002.
XX
XX 20-JUN-2001; 2001WO-US19692.
XX
XX 23-JUN-2000; 2000US-213637P.
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 10-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854288.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
XX
XX (GETH ) GENENTECH INC.
XX
XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
XX Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
XX Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX WPI; 2002-090516/12.
XX DR N-PSDB; ABL88203.
XX
XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
XX useful in diagnosis and treatment of cardiovascular (e.g. myocardial
XX infarction), endothelial or angiogenic disorders in a mammal .
XX
XX Claim 11; Fig 264; 565pp; English.
XX
XX ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
XX ABB85003. The PRO proteins and polynucleotides have cardiac, cytostatic,
XX antiangiogenic, hypotensive, vulnerary and antiarteriosclerotic

```

SQ	Sequence	273 AA;
----	----------	---------

3

55

Search completed: May 20, 2003, 17:05:07
Job time : 77 secs

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	178.5	16.0	1455	3	US-08-840-062-5	Sequence 5, Appli	
2	176.5	15.8	1479	3	US-08-840-062-4	Sequence 4, Appli	
3	174	13.6	1257	1	US-08-340-428B-49	Sequence 49, Appl	
4	174	15.6	2409	6	5180808-2	Patent No. 5180808	
5	170.5	15.3	1479	3	US-08-840-062-2	Sequence 2, Appli	
6	152	13.6	912	5	PCT-US95-03747-2	Sequence 2, Appli	
7	151	13.5	197	4	US-09-602-877A-99	Sequence 99, Appl	
8	149	13.4	174	1	US-07-641-971B-1	Sequence 1, Appli	
9	149	13.4	174	1	US-07-781-248A-1	Sequence 1, Appli	
10	149	13.4	320	1	US-08-365-103B-10	Sequence 10, Appl	
11	149	13.4	321	1	US-08-365-103B-8	Sequence 8, Appli	
12	148.5	13.3	652	2	US-08-751-305-2	Sequence 2, Appli	
13	144.5	13.0	404	4	US-09-517-605-2	Sequence 2, Appli	
14	143.5	12.9	372	2	US-08-513-278-4	Sequence 4, Appli	
15	143.5	12.9	372	6	5514582-4	Patent No. 5514582	
16	142	12.7	1487	3	US-08-840-062-7	Sequence 7, Appli	
17	141.5	12.7	125	3	US-08-722-126A-7	Sequence 7, Appli	
18	141.5	12.7	125	5	PCT-US95-04258-7	Sequence 7, Appli	
19	141.5	12.7	287	1	US-08-365-103B-4	Sequence 4, Appli	
20	141.5	12.7	300	1	US-08-365-103B-6	Sequence 6, Appli	
21	141.5	12.7	327	1	US-08-365-103B-2	Sequence 2, Appli	
22	138.5	12.4	117	6	5514582-7	Patent No. 5514582	
23	138.5	12.4	119	1	US-08-340-539A-12	Sequence 12, Appl	
24	137.5	12.3	110	6	5514582-12	Patent No. 5514582	
25	137.5	12.3	123	6	5514582-19	Patent No. 5514582	
26	137.5	12.3	492	4	US-09-724-86A-39	Sequence 39, Appl	
27	135	12.1	238	4	US-09-111-470-8	Sequence 8, Appli	

```

Db 865 ISHOKK-----FIMDGSKVDFVAVATGEPNPFANDENCVTMY-----TNSGF----- 907
QY 139 MFQWDDRCNMKNFICK---YSDEKPAVPSPREAGEETEELTPVLPETQOE----- 187
Db 908 ---WINDINGYNNFICORHSSINATAMP-----TPTTPGGCKEGWHLYKNK 953
QY 188 -----EDAKTKPESREAAALNL 204
Db 954 CFKIFGAFANEERKKSQDARQACKGL 978

RESULT 2
US-08-840-062-4
; Sequence 4, Application US/08840062
; Patent No. 611797
; GENERAL INFORMATION:
; APPLICANT: LASKY, LAURENCE A.
; APPLICANT: WU, KAI
; TITLE OF INVENTION: TYPE C LECTINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840.062
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1019R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1479 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-840-062-4

Query Match 15.8%; Score 176.5; DB 3; Length 1479;
Best Local Similarity 30.4%; Pred. No. 1.4e-09;
Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;

QY 7 SGQVPCRGGTORPCYKVIYFHDTSRLNFEFEAKACRRDGGOLVSISEDEOKLIEKFTIE 66
Db 395 SWQPF-----QHCYRL-----QAEKRSWQSKKACLRGGDLVSIHSMAELEFTIKQIK 434
QY 67 NLLPSDGDWGLRLRREEKQSNSTACQDLYAWTDGSIQFRNRYVDEPS---CGSEVCVW 123
Db 435 QEVE---ELWIGL-----NDLKLQNNFWSQDGLSVFTWHHPFEPNFRDSLEDCVT 483
QY 124 MYHQPAPAGGGPYMFQWDDRCNMKNFICKYSDEKPAVPSPREAG 171
Db 484 IW---GPEG-----RWNDSPCNQSLPSICKKAGQLSQGAABEDHG 520

RESULT 3
US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648465
; GENERAL INFORMATION:

```

```

; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Uwe
; APPLICANT: MARGOLIS, Renee K.
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340.428B
; FILING DATE: 14 No. 5648465ember 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/922,911
; FILING DATE: 03 August 1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: Margolis-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-340-428B-49

Query Match 15.6%; Score 174; DB 1; Length 1257;
Best Local Similarity 30.8%; Pred. No. 2e-09;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;

QY 17 QRPCYKVIYFHDTSRLNFEFEAKACRRDGGOLVSISEDEOKLIEKFTIENLLPSDGDW 76
Db 1037 QHCYR--YF---AHRRAWEADAERDCRRAGHLTSVHSPEEHKFINSF-----GHENSW 1085
QY 77 IGLRREEKQSNSTACQDLYAWTDGSIQFRNRYVDEPS---CGSEVCVWYHQPAPAG 133
Db 1086 IGLNDRTVRD-----FQWTDNTGLQYENWREKQDPNFFAGGDCVWYVAHENG--- 1134
QY 134 IGGPYMFQWDDRCNMKNFICK 156
Db 1135 -----RWNDVPCYNLPYVCK 1150

RESULT 4
5180808-2
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441.179
; FILING DATE: 27-NOV-1989
; SEQ ID NO:2:
; LENGTH: 2409
5180808-2

```

```

RESULT 6
PCT-US95-03747-2
; Sequence 2, Application PC/TUS9503747
; GENERAL INFORMATION:
; APPLICANT: LA JOLLA CANCER RESEARCH FOUNDATION
; TITLE OF INVENTION: Brevican, A Glial Cell Proteoglycan
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/03747
; FILING DATE: 27-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Imbra, Richard J.
; REGISTRATION NUMBER: 37,643
; REFERENCE/DOCKET NUMBER: FP-LJ 1453
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-03747-2

Query Match 13.6%; Score 152; DB 5; Length 912;
Best Local Similarity 28.8%; Pred.No.2.6e-07;
Matches 47; Conservative 17; Mismatches 61; Indels 38; Gaps

QY 17 QRCYKVIYPHDTSRLNLFEEAKEACRRDGGQLVSTSEDEOKLIEFIENLLPDSGDWF 76
Db 697 QGACYK----HFSARR-SWEEAENKCRMYCAHLASISTPEEQLFINNRYEQ-----W 745

QY 77 IGLRRREKOSNSTACQDLVATDGTSGISQFNKYYVDEPS---CGSEVCVVM-YHQPSAPA 132
Db 746 IGL-----NDRITGDFLWSDGVLPVLLYNNNGPDSYFLSGENCVVMVMDQG--- 794

QY 133 GIGGPFYMFQNDRCNNKNFNICKYS----DEKPAVPSRAEG 171
Db 795 -----QMSDVPNCVHLSTCKMGLVSCGPPPELPAEVEFG 829

RESULT 7
US-09-602-877A-99
; Sequence 99, Application US/09602877A
; Patent No. 6432707
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.446C5
; CURRENT APPLICATION NUMBER: US/09/602,877A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 99

```

	Best Local Similarity	30.2a; Pred. No. 5.7e-08;			
	Matches	55; Conservative	20; Mismatches	57; Indels	50; Gaps
Qy	17	QRPCYKVIYHDTGRRINLFEEAKACRRDGGQLVSISEDEQKLIEKFNIENLLPSGDWF	76		
		: : : : :			
Dd	24	QRKYC--YFGKGTKQ--WYHARYACDDMEGLVSIHSPEEQDFLTKH----	73		
		: : : : :			
Qy	77	IGLRREKOSNSTACODLVAWTDGISIQSFNNMYVDSPSCGS--EVCVMVHQPSAPAGI	134		
		: : : :			
Dd	74	IGLRNLDLKE-----FIWDGSHVDYSNNAQCEPTSRSGQEDCVMM-----	RGS	118	
		: : : :			
Qy	135	GGPYMFQWNDRCMNK--NNFICKYSDEKPAY---PSPREAEGE-----	ETELTPPV	180	
		: : : :			
Dd	119	G-----RWNDAFCDRLKGAWYC----DKLATCTFPASGESAESMGDPDRPDGRUPTPS	169		
		: : : :			
Qy	181	LP	182		
Dd	170	AP	171		

RESULT 9
 US-07-781-248A-1
 ; Sequence 1, Application US/07781248A
 ; Patent No. 5246699
 ; GENERAL INFORMATION:
 ; APPLICANT: Debre, Patrice
 ; APPLICANT: Mossalayi, Mohammed D
 ; TITLE OF INVENTION: MATURATION OF HEMATOPOIETIC CELLS
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Irving M. Fishman, CIBA-GEIGY Corporation
 ; STREET: 556 Morris Avenue
 ; CITY: Summit
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07901
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07781,248A
 ; FILING DATE: 19911230
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 90103565
 ; FILING DATE: 09-MAY-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ikeler, Barbara J.
 ; REGISTRATION NUMBER: 36,170
 ; REFERENCE/DOCKET NUMBER: 4-18065/A/DEB
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 908-277-3368
 ; TELEFAX: 908-277-4306
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 174 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; ORIGINAL SOURCE:
 ; CELL TYPE: Human B. Cells
 ; CELL LINE: CHO cells transformed with pCAL8-BF-ND
 US-07-781-248A-1

Query Match 13.4%; Score 149; DB 1; Length 174;
Best Local Similarity 30.2%; Pred. No. 5.7e-08;
Matches 55; Conservative 20; Mismatches 57; Indels 50; Gaps

Db 266 G-----RWNDAFCDRLGAWC-----DRLATCTPPASEGSAESMGDPDRDPDGRLLPTFS 316

QY 181 LP 182

Db 317 AP 318

RESULT 12
US-08-751-305-2
; Sequence 2, Application US/08751305

```
; Patent No. 5965439
; GENERAL INFORMATION:
; APPLICANT: Tenner et al., Andrea J.
; TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,305
; FILING DATE: 18-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07306/012001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-751-305-2

Query Match      13.3%; Score 148.5; DB 2; Length 652;
Best Local Similarity 24.4%; Pred. No. 3.9e-07;
Matches 50; Conservative 37; Mismatches 75; Indels 43; Gaps 9;

QY 5 LLSGQP-----VCRGQTQPCYKVIYHDTSRRLNFEAEKACRRDGGOLVSIE 53
   || || || || || || || || || || || || || || || || || || ||
Db 13 LLLTPGAGTGADTEAVVCGV---TACYTA-----HSGKLSAAEAQHCNONGNLATVK 64

QY 54 SEDQKLIKIEIENLLPSD-----GDFWIGLRRREEKQSNSTACQDLYAWT-DGSISQ 105
   || || || || || || || || || || || || || || || || || || ||
Db 65 SKEEAHVQRYVLAQLLRREALTARMSKFWIGLQREKGCILDPSLPLKGFWSWVGGEDTP 124

QY 106 FRNMYVD-EPSCGSEVCVVM---YHOPAPAGIGGYPYFQWDDRCNWKNN-----FI 154
   || || || || || || || || || || || || || || || || || || ||
Db 125 YSNHKLERNSCISKRCVSLLLDLSQPLLPNRLP-----KWSEGPCGSPGSPGNIIEGFV 179

QY 155 KYSDEKPAVPSREAEGETELTTP 179
   || || || || || || || || || || || || || || || || || || ||
Db 180 CKFSEKMGCRPLALGGPGQVITYTP 204

RESULT 13
US-09-517-605-2
; Sequence 2, Application US/09517605
; Patent No. 6391567
; GENERAL INFORMATION:
; APPLICANT: Littman, Dan R.
; APPLICANT: Kwon, Douglas S.
; APPLICANT: van Kooyk, Yvette
; APPLICANT: Geljtenbeck, Tneo
; TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY INTO
; FILE REFERENCE: 1049-1-017
; CURRENT APPLICATION NUMBER: US/09/517,605
; CURRENT FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 2
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-517-605-2

Query Match      13.0%; Score 144.5; DB 4; Length 404;
Best Local Similarity 29.7%; Pred. No. 5.3e-07;
Matches 47; Conservative 23; Mismatches 47; Indels 41; Gaps 10;

QY 25 YFHTSRRLNFEAEKACRRDGGOLVSIESEDEQKLIKIEFIENLLPSDGFHWIGLRRREE 84
   || || || || || || || || || || || || || || || || || || ||
Db 268 YFMSNSQR-NMHDSTITACKVEGAQLVWIKSAEQNFLOLQSSR---SNRFTWGLSLDLNQ 323

QY 85 KOSNSTACQDLYAWTDGS--ISOFRN-WYVDEP-SCGSEVCVVMYHOPAPAGIGGYPYMF 140
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 -----EGTWQVVDGSPFLPSPFKQYWNRGEPNNGVEDCAEF-----SGNG----- 363

QY 141 QWDDRCNMKNFICKYS-----DEK-----PAVPS 166
   ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 364 -WDDKCNLAKFWICKSAACSRDEQFLSPAPATPN 400

RESULT 14
US-08-513-278-4
; Sequence 4, Application US/08513278
; Patent No. 5840844
; GENERAL INFORMATION:
; APPLICANT: LASKY, LAURENCE A.
; APPLICANT: STACHELL, SCOTT E.
; APPLICANT: ROSEN, STEVEN D.
; APPLICANT: SINGER, MARK S.
; APPLICANT: YEDNOCK, TED A.
; TITLE OF INVENTION: LYMPHOCYTE HOMING RECEPTORS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/513,278
; FILING DATE: 10-AUG-1995
; CLASSIFICATION: 5530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/059027
; FILING DATE: 06-MAY-1993
; APPLICATION NUMBER: 07/786149
; FILING DATE: 31-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/315015
; FILING DATE: 23-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 565D1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-513-278-4
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:06:10 ; Search time 56 Seconds
(without alignments)
364.814 Million cell updates/sec

Title: US-09-887-855-5
Perfect score: 1115
Sequence: 1 ATGRLSGQVPCRGQTQPC.....EEDAKTFKESREALNLAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PublishedApplications_AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pap.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1115	100.0	206	10	US-09-887-855-5
2	1115	100.0	374	9	US-10-149-819-15
3	1115	100.0	374	10	US-09-887-855-2
4	1101	98.7	382	9	US-09-905-291A-137
5	1101	98.7	382	9	US-09-902-853-137
6	1101	98.7	382	9	US-09-907-824-137
7	1101	98.7	382	9	US-09-907-841-137
8	1101	98.7	382	9	US-09-904-011-137
9	1101	98.7	382	9	US-10-174-590-20
10	1101	98.7	382	9	US-10-176-758-20
11	1101	98.7	382	9	US-10-175-737-20
12	1101	98.7	382	9	US-09-906-742-137
13	1101	98.7	382	9	US-10-173-706-20
14	1101	98.7	382	9	US-10-175-738-20
15	1101	98.7	382	9	US-10-175-752-20
16	1101	98.7	382	9	US-10-176-482-20
17	1101	98.7	382	9	US-10-176-757-20
18	1101	98.7	382	9	US-10-176-913-20
19	1101	98.7	382	9	US-10-180-552-20

20	1101	98.7	382	9	US-10-180-557-20	Sequence 20, Appl
21	1101	98.7	382	9	US-09-906-838-137	Sequence 137, App
22	1101	98.7	382	9	US-09-907-613-137	Sequence 137, App
23	1101	98.7	382	9	US-09-907-942-137	Sequence 137, App
24	1101	98.7	382	9	US-10-173-700-20	Sequence 20, Appl
25	1101	98.7	382	9	US-10-174-572-20	Sequence 20, Appl
26	1101	98.7	382	9	US-10-174-579-20	Sequence 20, Appl
27	1101	98.7	382	9	US-10-174-582-20	Sequence 20, Appl
28	1101	98.7	382	9	US-10-174-588-20	Sequence 20, Appl
29	1101	98.7	382	9	US-10-175-739-20	Sequence 20, Appl
30	1101	98.7	382	9	US-10-175-740-20	Sequence 20, Appl
31	1101	98.7	382	9	US-10-175-743-20	Sequence 20, Appl
32	1101	98.7	382	9	US-10-176-488-20	Sequence 20, Appl
33	1101	98.7	382	9	US-10-176-492-20	Sequence 20, Appl
34	1101	98.7	382	9	US-10-176-747-20	Sequence 20, Appl
35	1101	98.7	382	9	US-10-176-750-20	Sequence 20, Appl
36	1101	98.7	382	9	US-10-176-985-20	Sequence 20, Appl
37	1101	98.7	382	9	US-10-176-987-20	Sequence 20, Appl
38	1101	98.7	382	9	US-10-176-991-20	Sequence 20, Appl
39	1101	98.7	382	9	US-10-176-993-20	Sequence 20, Appl
40	1101	98.7	382	9	US-10-176-993-20	Sequence 20, Appl
41	1101	98.7	382	9	US-10-184-658-20	Sequence 20, Appl
42	1101	98.7	382	9	US-10-173-695-20	Sequence 20, Appl
43	1101	98.7	382	9	US-10-173-697-20	Sequence 20, Appl
44	1101	98.7	382	9	US-10-173-705-20	Sequence 20, Appl
45	1101	98.7	382	9	US-10-174-576-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-09-887-855-5
; Sequence 5, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dick M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-887-855-5

Query Match	100.0%	Score 1115;	DB 10;	Length 206;
Best Local Similarity	100.0%	Pred. No. 6.6e-103;		
Matches 206;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGRLSGQVPCRGQTQPCYKVIYFHDTSRLNFEAEACRDRDGGOLVSI	ESDEOKL 60	
Db	1	ATGRLSGQVPCRGQTQPCYKVIYFHDTSRLNFEAEACRDRDGGOLVSI	ESDEOKL 60	
Qy	61	IEKFTENLLPSGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEP	SCGSEV 120	
Db	61	IEKFTENLLPSGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEP	SCGSEV 120	
Qy	121	CVVMYHQPASAGIGGPPYMFQWDDRCNKNFNICKYSDERPAVPSRAEGE	TELTPV 180	
Db	121	CVVMYHQPASAGIGGPPYMFQWDDRCNKNFNICKYSDERPAVPSRAEGE	TELTPV 180	
Qy	181	LPEETQEDAKTKFKESREALNLAY	206	
Db	181	LPEETQEDAKTKFKESREALNLAY	206	

RESULT 2
US-10-149-819-15

Matches	206;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps
QY	1	ATGRLSSGPVCRGGTORPCVKVIYFDHTSRRLLNFEAEKACRRDGGQLVSTESDEQKL	60					
Db	22	ATGRLSSGPVCRGGTORPCVKVIYFDHTSRRLLNFEAEKACRRDGGQLVSTESDEQKL	81					
QY	61	IEKFIEIALLPSDGDWFVIGLRRRREEKQSNTACODLYAWTDGSIOSFRNMYYVDPEPSCGSEV	120					
Db	82	IEKFIEIALLPSDGDWFVIGLRRRREEKQSNTACODLYAWTDGSIOSFRNMYYVDPEPSCGSEV	141					
QY	121	CVVMYHQPSPAGIGGPFYFQWDDRCNNMKNNFICKYSEKPAVPSPREAGEGETELTTPV	180					
Db	142	CVVMYHQPSPAGIGGPFYFQWDDRCNNMKNNFICKYSEKPAVPSPREAGEGETELTTPV	201					
QY	181	LPEETQEEDAKKTFFKESREAALNLAY	206					
Db	202	LPEETQEEDAKKTFFKESREAALNLAY	227					
RESULT 4								
US-09-905-291A-137								
; Sequence 137, Application US/09905291A								
; Patent No. US20020160374A1								
; GENERAL INFORMATION:								
; APPLICANT: Genentech, Inc.								
; APPLICANT: Ashkenazi, Avi								
; APPLICANT: Botstein, David								
; APPLICANT: Desnoyers, Luc								
; APPLICANT: Eaton, Dan L.								
; APPLICANT: Ferrara, Napoleone								
; APPLICANT: Filvaroff, Ellen								
; APPLICANT: Fong, Sherman								
; APPLICANT: Gao, Wei-Qiang								
; APPLICANT: Gerber, Hanspeter								
; APPLICANT: Gerritsen, Mary E.								
; APPLICANT: Goddard, A.								
; APPLICANT: Godowski, Paul J.								
; APPLICANT: Grimaldi, Christopher J.								
; APPLICANT: Gurney, Austin L.								
; APPLICANT: Hillan, Kenneth, J.								
; APPLICANT: Kljavin, Ivar J.								
; APPLICANT: Mather, Jennie P.								
; APPLICANT: Pan, James								
; APPLICANT: Paoni, Nicholas F.								
; APPLICANT: Roy, Margaret Ann								
; APPLICANT: Stewart, Timothy A.								
; APPLICANT: Tumas, Daniel								
; APPLICANT: Williams, P. Mickey								
; APPLICANT: Wood, William, I.								
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same								
; FILE REFERENCE: 10466-14								
; CURRENT APPLICATION NUMBER: US/09/905,291A								
; CURRENT FILING DATE: 2001-07-12								
; PRIOR APPLICATION NUMBER: PCT/US00/04414								
; PRIOR FILING DATE: 2000-02-22								
; PRIOR APPLICATION NUMBER: US 60/143,048								
; PRIOR FILING DATE: 1999-07-07								
; PRIOR APPLICATION NUMBER: US 60/145,698								
; PRIOR FILING DATE: 1999-07-26								
; PRIOR APPLICATION NUMBER: US 60/146,222								
; PRIOR FILING DATE: 1999-07-28								
; PRIOR APPLICATION NUMBER: PCT/US99/20594								
; PRIOR FILING DATE: 1999-09-08								
; PRIOR APPLICATION NUMBER: PCT/US99/20944								
; PRIOR FILING DATE: 1999-09-13								
; PRIOR APPLICATION NUMBER: PCT/US99/21090								
; PRIOR FILING DATE: 1999-09-15								
; PRIOR APPLICATION NUMBER: PCT/US99/21547								
; PRIOR FILING DATE: 1999-09-15								
; PRIOR APPLICATION NUMBER: PCT/US99/23089								
; PRIOR FILING DATE: 1999-10-05								
; PRIOR APPLICATION NUMBER: PCT/US99/23214								

```
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-291A-137

Query Match      98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLYSI 52
Db 22 ATGRLLSASDLRLRGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLYSI 81
Qy 53 ESEDEQKLIKFIENLLPSDGFHWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVD 112
Db 82 ESEDEQKLIKFIENLLPSDGFHWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVD 141
Qy 113 EPSCGSEVCVVMYHOPSPAGIGGYPMFQWDDRCNMKNFICKYSDEKPAVPSREAGE 172
Db 142 EPSCGSEVCVVMYHOPSPAGIGGYPMFQWDDRCNMKNFICKYSDEKPAVPSREAGE 201
Qy 173 ETELTTPVLPETQEDAKKTFKESREAAALNLAY 206
Db 202 ETELTTPVLPETQEDAKKTFKESREAAALNLAY 235

RESULT 5
US-09-902-853-137
; Publication No. US20020192659A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Oiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
```

```
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,853
; PRIOR APPLICATION NUMBER: US/09/665,350
; PRIOR FILING DATE: 2001-07-10
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-902-853-137

Query Match      98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLYSI 52
Db 22 ATGRLLSASDLRLRGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLYSI 81
Qy 53 ESEDEQKLIKFIENLLPSDGFHWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVD 112
Db 82 ESEDEQKLIKFIENLLPSDGFHWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVD 141
Qy 113 EPSCGSEVCVVMYHOPSPAGIGGYPMFQWDDRCNMKNFICKYSDEKPAVPSREAGE 172
Db 142 EPSCGSEVCVVMYHOPSPAGIGGYPMFQWDDRCNMKNFICKYSDEKPAVPSREAGE 201
Qy 173 ETELTTPVLPETQEDAKKTFKESREAAALNLAY 206
Db 202 ETELTTPVLPETQEDAKKTFKESREAAALNLAY 235

RESULT 6
US-09-907-824-137
; Sequence 137, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
```



```
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-841-137

Query Match          98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 52
    |||||||
Db 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 81
    |||||||

Qy 53 ESEDEQKLIKFIEIENLPSDGFWTGLRRREEKQSNSTACODLYAWTDGSIQFRNMYVD 112
    |||||||
Db 82 ESEDEQKLIKFIEIENLPSDGFWTGLRRREEKQSNSTACODLYAWTDGSIQFRNMYVD 141
    |||||||

Qy 113 EPSCGSEVCVVMYHOPSPAGIGGPFQWDDRCNMKNNFICKYSDEKPAVPSREAGE 172
    |||||||
Db 142 EPSCGSEVCVVMYHOPSPAGIGGPFQWDDRCNMKNNFICKYSDEKPAVPSREAGE 201
    |||||||

Qy 173 ETELTTPVLPETQEDAKKTFKESREAAALNLAY 206
    |||||||
Db 202 ETELTTPVLPETQEDAKKTFKESREAAALNLAY 235
    |||||||

RESULT 8
US-09-904-011-137
; Sequence 137, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
```

```
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-904-011-137

Query Match          98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 52
    |||||||
Db 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 81
    |||||||

Qy 53 ESEDEQKLIKFIEIENLPSDGFWTGLRRREEKQSNSTACODLYAWTDGSIQFRNMYVD 112
    |||||||
Db 82 ESEDEQKLIKFIEIENLPSDGFWTGLRRREEKQSNSTACODLYAWTDGSIQFRNMYVD 141
    |||||||

Qy 113 EPSCGSEVCVVMYHOPSPAGIGGPFQWDDRCNMKNNFICKYSDEKPAVPSREAGE 172
    |||||||
Db 142 EPSCGSEVCVVMYHOPSPAGIGGPFQWDDRCNMKNNFICKYSDEKPAVPSREAGE 201
    |||||||

Qy 173 ETELTTPVLPETQEDAKKTFKESREAAALNLAY 206
    |||||||
Db 202 ETELTTPVLPETQEDAKKTFKESREAAALNLAY 235
    |||||||

RESULT 9
US-10-174-590-20
; Sequence 20, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
```


	Query Match	98.7%	Score 1101;	DB 9;	Length 382;
	Best Local Similarity	96.3%	Pred. No. 3.6e-101;		
	Matches 206;	Conservative	0;	Mismatches 0;	Indels 8; Gaps 1;
QY	1	ATGRLLS-----GQPCVCGGTQRCYKVIYFHDTSRRLNFEAEACRBDGGQLVSI	52		
Db	22	ATGRLLSASDLRGQPCVCGGTQRCYKVIYFHDTSRRLNFEAEACRBDGGQLVSI	81		
QY	53	ESDEOKLIKBFITENLPDSQGFWMIGLRRREEKOSNSTACODLYAWTDGSIQSPRNMYVD	112		

: GENERAL INFORMATION:
 : APPLICANT: Genentech, Inc.
 : APPLICANT: Ashkenazi, Avi
 : APPLICANT: Botstein, David
 : APPLICANT: Desnovers, Luc
 : APPLICANT: Eaton, Dan L.
 : APPLICANT: Ferrara, Napoleone
 : APPLICANT: Fivvaroff, Ellen

```

: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, A.
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth, J.
: APPLICANT: Kijavlin, Ivar J.
: APPLICANT: Mather, Jennie P.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: 10466-14
: CURRENT APPLICATION NUMBER: US/09/906,742
: CURRENT FILING DATE: 2001-07-16
: PRIOR APPLICATION NUMBER: 09/665,350
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: PCT/US00/04414
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: PCT/US99/20944
: PRIOR FILING DATE: 1999-09-13
: PRIOR APPLICATION NUMBER: PCT/US99/21090
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/21547
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/23089
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: PCT/US99/28214
: PRIOR FILING DATE: 1999-11-29
: PRIOR APPLICATION NUMBER: PCT/US99/28313
: PRIOR FILING DATE: 1999-11-30
: PRIOR APPLICATION NUMBER: PCT/US99/28564
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/28565
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30095
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: PCT/US99/30911
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US99/30999
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 423
: SEQ ID NO 137
: LENGTH: 382
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-09-906-742-137

```

```

Query Match      98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLIS-----GQVCRGGTQRCYKVIYPHDTSRRLNFEEAKERRDGGQLVSI 52
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 22 ATGRLISASLDLGGGQVCRGGTQRCYKVIYPHDTSRRLNFEEAKERRDGGQLVSI 81
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 53 ESEDEQKLIERFIENLLPSDGDFFWIGLRREEKQSNSTACODLAWTDDGSIQFRNMYVD 112
      |||||||
Db 82 ESEDEQKLIERFIENLLPSDGDFFWIGLRREEKQSNSTACODLAWTDDGSIQFRNMYVD 141
      |||||||
QY 113 EPSGSEVCVVMYHOPSPAGIGGPGYFQWDDRCNMKNNFICKYSDEKPAVPSREAGE 172
      |||||||
Db 142 EPSGSEVCVVMYHOPSPAGIGGPGYFQWDDRCNMKNNFICKYSDEKPAVPSREAGE 201
      |||||||
QY 173 ETELTTPVLPETQBEDAKKTFKESREAALNLAY 206
      |||||||
Db 202 ETELTTPVLPETQBEDAKKTFKESREAALNLAY 235
      |||||||

RESULT 13
US-10-173-706-20
; Sequence 20, Application US/10173706
; Publication No. US2003002293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEOT
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 20
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-20

```

Query Match 98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. NO. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy	1	ATGRLLS-----GQPVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI	52
Db	22	ATGRLLSASDLRLGGQPVCRRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI	81
Qy	53	ESEDEQKLI EKFIENLLPSDGDWFI GLRRREEQKSNSTACODLIYAWTDGSI SFRNMYVD	112
Db	82	ESEDEQKLI EKFIENLLPSDGDWFI GLRRREEQKSNSTACODLIYAWTDGSI SFRNMYVD	141
Qy	113	EPSCGSEVCVVMYHQPAPAGIGPYPFQWDDRCNNKNNFICKYSDEKPAVPSREAEGE	172
Db	142	EPSCGSEVCVVMYHQPAPAGIGPYPFQWDDRCNNKNNFICKYSDEKPAVPSREAEGE	201
Qy	173	ETELTTVPVLPETQOEEDAKKTFKESREAALNLAY	206
Db	202	ETELTTVPVLPETQOEEDAKKTFKESREAALNLAY	235

```

RESULT 14
US-10-175-738-20
; Sequence 20, Application US/10175738
; Publication No. US20030022594A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
;

```

```
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 20
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-20

Query Match      98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQPVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLDLRGGQPVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 81
Qy 53 ESEDEQKLIKFIENLLPSDGDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD 112
Db 82 ESEDEQKLIKFIENLLPSDGDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD 141
Qy 113 EPSCGSEVCVVMYHQPSPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREARGE 172
Db 142 EPSCGSEVCVVMYHQPSPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREARGE 201
Qy 173 ETELTPVLPETQEDAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPETQEDAKKTFKESREAAALNLAY 235
```

Search completed: May 20, 2003, 17:15:30
Job time : 58 secs

```
RESULT 15
US-10-175-752-20
; Sequence 20, Application US/10175752
; Publication No. US2003002295A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 20
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-20

Query Match      98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:03:25 ; Search time 44 Seconds
(without alignments)
450.084 Million cell updates/sec

Title: US-09-887-855-5

Perfect score: 1115

Sequence: 1 ATGLLSGQPVCRGQTORPC.....EEDAKKTFKESREAAALNLAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	182	16.3	1456	1	A36563	mannose receptor p
2	178.5	16.0	1455	1	A48925	mannose receptor p
3	177	15.9	1268	2	S52781	neurocan - mouse
4	174.5	15.7	1643	2	T14274	versican precursor
5	174.5	15.7	3381	2	T42389	versican precursor
6	174	15.6	1257	2	S28764	neurocan precursor
7	174	15.6	2397	1	A55535	versican precursor
8	174	15.6	2409	1	A60979	versican precursor
9	171	15.3	3562	2	A47171	chondroitin sulfat
10	170.5	15.3	1479	2	T42710	mannose receptor,
11	158.5	14.2	1340	2	A39808	proteoglycan core
12	158.5	14.2	2327	2	T42630	aggreacan precursor
13	158.5	14.2	2415	1	A39086	E-selectin precurs
14	154.5	13.9	612	2	B42755	proteoglycan core
15	153.5	13.8	2124	2	A28452	aggreacan precursor
16	152	13.6	912	2	A54423	proteoglycan core
17	149.5	13.4	459	2	T24425	brevican precursor
18	149	13.4	321	1	LNHUP6	hypothetical prote
19	148.5	13.3	330	2	T46256	IgE Fc receptor II
20	148.5	13.3	2132	1	A55182	brevican - human (
21	148	13.3	253	2	E89130	aggreacan precursor
22	147	13.2	883	2	S57653	protein F52E1.2 [1
23	146.5	13.1	162	1	LNRC1	brevican precursor
24	146	13.1	2109	1	I50421	lectin BRA3-1 prec
25	145.5	13.0	742	2	JC7595	aggreacan precursor
26	145	13.0	883	2	S49126	scavenger receptor
27	144.5	13.0	173	2	S10548	brevican precursor
28	144.5	13.0	372	2	S23936	lectin - barnacle
29	144.5	13.0	404	2	A46274	L-selectin precurs
						HIV gp120-binding

30 143.5 12.9 129 2 JC4329
31 143.5 12.9 372 1 A32375
32 143.5 12.9 463 2 T26655
33 142.5 12.8 131 2 JC5058
34 142 12.7 1487 2 S48719
35 141.5 12.7 331 1 LNMSER
36 140.5 12.6 162 1 LNRC3
37 140 12.6 370 2 S22124
38 139 12.5 248 1 LNHP6
39 139 12.5 248 1 LNHP6
40 139 12.5 248 1 LNHP6
41 139 12.5 283 1 LNHP6
42 138.5 12.4 152 2 JC4690
43 138.5 12.4 202 2 JC4031
44 138 12.4 280 2 T29200
45 137.5 12.3 309 1 S34198

ALIGNMENTS

RESULT 1

A36563
mannose receptor precursor - human

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A36563; A60926; A44255; B44255; C44255; D44255; E44255; F44255; G44255; H44255; I44255; J44255; K44255; L44255; M44255; N44255; O44255; P44255; Q44255; R44255; S44255; T44255; U44255; V44255; W44255; X44255; Y44255; Z44255

R:Taylor, M.E.; Conary, J.T.; Lennartz, M.R.; Stahl, P.D.; Drickamer, K.

J. Biol. Chem. 265, 12156-12162, 1990

A:Title: Primary structure of the mannose receptor contains multiple motifs resembling A:Reference number: A36563; MUID:90324192; PMID:2373685

A:Accession: A36563

A:Molecule type: mRNA

A:Residues: 1-1456 <TAY>

A:Cross-references: GB:J05550; MID:gl88675; PIDN:AAA59868.1; PID:gl88676

A:Note: parts of this sequence, including the amino end of the mature protein, were

R:Ezekowitz, R.A.B.; Sastry, K.; Bailly, P.; Warner, A.

J. Exp. Med. 172, 1785-1794, 1990

A:Title: Molecular characterization of the human macrophage mannose receptor: demon

A:Reference number: A60926; MUID:91079783; PMID:2258707

A:Accession: A60926

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1333, 'T', 1335-1456 <EZE>

A:Cross-references: GB:X55635

A:Note: translation of the nucleotide sequence is incomplete

A:Note: in the authors' translation additional residues Pro-Glu-Ile are shown after R:Kim, S.-J.; Ruiz, N.; Bezouska, K.; Drickamer, K.

Genomics 14, 721-727, 1992

A:Title: Organization of the gene encoding the human macrophage mannose receptor (M

A:Reference number: A44255; MUID:93052405; PMID:1294118

A:Accession: A44255

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: DNA

A:Residues: 155-233, 'KSAL', 238-283, 346-428, 492-569, 631-714, 716-719, 783-820, 'N', 822-1

A:Note: sequence extracted from NCBI backbone (NCBIP:118415, NCBIP:118421, NCBIP:118

C:Genetics:

A:Gene: GDB:MRCL

A:Cross-references: GDB:133759; OMIM:153618

A:Map position: 10p13-10p13

C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type

C:Keywords: duplication; lectin; tandem repeat; transmembrane protein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:168-209/Domain: fibronectin type II repeat homology <2F1>

F:223-340/Domain: C-type lectin homology <LCH1>

F:362-486/Domain: C-type lectin homology <LCH2>

F:945-1079/Domain: C-type lectin homology <LCH3>

Query Match

Best Local Similarity 16.3%; Score 182; DB 1; Length 1456;

Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;

OY 21 YKVIYFHTDTSRLNPFNAKEACRRDGGQLVSTSEDEQKLEKFTENLLPSDGFWIGLR 80

00b 908 ---WNDINGYPNFIQQRHSSINATAMP-----TPTTPGGCKEGWHLK 953

QY 76 WIGLRREKQSNSTACQDLYAWTSGISQFRNYYVDEP-----SCGSEVCVVMYHQPSPAP 131
 DB 1472 WIGL-----NDKMFHDFRMTDGTSLQYENWRPNQDSFSTGDCVVIWHENG-- 1521
 QY 132 AGIGPYMFQWDDRCNKNFNICKYS-----DEKPAVPSREAEGE----- 172
 DB 1522 -----QWMDVPCNHYLTTCCKGTVACQPPVVENAKTFGKMKPRYENSLIRYHC 1572
 QY 173 -----ETELTT-----PVL-----PEETQEDAKTKFKESREAAALN 203
 DB 1573 KDGFIQRHLPTIRCLNGRWAMPKICTLNPISAYQRTYSKKYFNKSSSAKDN 1623

RESULT 5
 T42389
 N:Alternate names: chondroitin sulfate proteoglycan
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000
 C:Accession: T42389
 R:Schmalheid, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
 J. Biol. Chem. 273, 15758-15764, 1998
 A:Title: Versican V2 is a major extracellular matrix component of the mature bovine brain
 A:Reference number: Z17954; MUID:98288320; PMID:9624174
 A:Accession: T42389
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3381 <SCH>
 A:Cross-references: EMBL:AF060456; NID:g3253299; PID:g3253300; PIDN:AAC24358.1
 C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin
 C:Keywords: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein
 F:1-20/Domain: signal sequence #status predicted <Sig>
 F:21-3381/Product: versican, splice form V0 #status predicted <MAT>
 F:257,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356,26

Query Match 15.7%; Score 174.5; DB 2; Length 3381;
 Best Local Similarity 25.5%; Pred. No. 1.1e-06;
 Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;

QY 17 QRPCKYKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDEQKLIKFIENLLPSDGDGF- 75
 DB 3162 QGQCYK--YF---AHRRTWDAARECRQLQGAHLTSLSHEEQMFVNRV-----CHDYQ 3209

QY 76 WIGLRREKQSNSTACQDLYAWTSGISQFRNYYVDEP-----SCGSEVCVVMYHQPSPAP 131
 DB 3210 WIGL-----NDKMFHDFRMTDGTSLQYENWRPNQDSFSTGDCVVIWHENG-- 3259

QY 132 AGIGPYMFQWDDRCNKNFNICKYS-----DEKPAVPSREAEGE----- 172
 DB 3260 -----QWMDVPCNHYLTTCCKGTVACQPPVVENAKTFGKMKPRYENSLIRYHC 3310

QY 173 -----ETELTT-----PVL-----PEETQEDAKTKFKESREAAALN 203
 DB 3311 KDGFIQRHLPTIRCLNGRWAMPKICTLNPISAYQRTYSKKYFNKSSSAKDN 3361

RESULT 6
 S28764
 neurocan precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
 C:Accession: S28764
 R:Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
 J. Biol. Chem. 267, 19536-19547, 1992
 A:Title: Cloning and primary structure of neurocan, a developmentally regulated, aggrega
 A:Reference number: S28764; MUID:92406907; PMID:1326557
 A:Accession: S28764
 A:Molecule type: mRNA
 A:Residues: 1-1257 <RAU>
 A:Cross-references: EMBL:M97161; NID:g205649; PIDN:AAC37679.1; PID:g205650
 C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EC
 C:Keywords: chondroitin sulfate proteoglycan; glycoprotein

F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-1257/Product: neurocan #status predicted <MAT>
 F:176-253/Domain: link protein repeat homology <LNK1>
 F:274-355/Domain: link protein repeat homology <LNK2>
 F:364-366/Region: cell attachment (R-G-D) motif
 F:953-984/Domain: EGF homology <EGF>
 F:1029-1149/Domain: C-type lectin homology <LCH>
 F:1150-1212/Domain: complement factor H repeat homology <FHD>
 F:121,339,737,967,1164/Binding site: carboxydrate (Asn) (covalent) #status predicted
 F:372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
 F:944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 15.6%; Score 174; DB 2; Length 1257;
 Best Local Similarity 30.8%; Pred. No. 4e-07;
 Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;

QY 17 QRPCKYKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDEQKLIKFIENLLPSDGDGF 76
 DB 1037 QGHCYR--YF---AHRRTWDAARECRRLRAGHLTSHVSPHEHAFINSF-----GHNSW 1085

QY 77 IGLRRREKQSNSTACQDLYAWTSGISQFRNYYVDEPS---CGSEVCVVMYHQPSPAP 133
 DB 1086 IGLNDRTVERD-----FQMDTNGTGLQYENWRPNQDSFSTGDCVVMVAHENG--- 1134

QY 134 IGGPYMFQWDDRCNKNFNICK 156
 DB 1135 -----RWMDVPCNHYLTTCCKGTVACQPPVVENAKTFGK 2312

RESULT 7
 A55535
 N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan gen
 N:Contains: glial hyaluronate-binding protein
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A55535
 R:Itto, K.; Shinomura, T.; Zako, M.; Ujita, M.; Kimata, K.
 J. Biol. Chem. 270, 958-965, 1995
 A:Title: Multiple forms of mouse PG-M, a large chondroitin sulfate proteoglycan gen
 A:Reference number: A55535; MUID:95122551; PMID:7822336
 A:Accession: A55535
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2397 <RES>
 A:Cross-references: GB:D16263; NID:g862460; PIDN:BAAO1796.1; PID:g862461
 C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology
 F:1-20/Domain: signal sequence #status predicted <Sig>
 F:21-1654/Domain: versican #status predicted <MAT>
 F:167-244/Domain: link protein repeat homology <LNK1>
 F:265-346/Domain: link protein repeat homology <LNK2>
 F:2095-2126/Domain: EGF homology <EG1>
 F:2133-2164/Domain: EGF homology <EG2>
 F:2171-2291/Domain: C-type lectin homology <LCH>
 F:2298-2354/Domain: complement factor H repeat homology <FHD>

Query Match 15.6%; Score 174; DB 1; Length 2397;
 Best Local Similarity 28.5%; Pred. No. 8.4e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

QY 17 QRPCKYKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDEQKLIKFIENLLPSDGDGF- 75
 DB 2179 QGQCYK--YF---AHRRTWDAARECRQLQGAHLTSLSHEEQMFVNRV-----CHDYQ 2226

QY 76 WIGLRREKQSNSTACQDLYAWTSGISQFRNYYVDEP-----SCGSEVCVVMYHQPSPAP 131
 DB 2227 WIGL-----NDKMFHDFRMTDGTSLQYENWRPNQDSFSTGDCVVIWHENG-- 2276

QY 132 AGIGPYMFQWDDRCNKNFNICKYS-----DEKPAVPSREAEGE 172
 DB 2277 -----QWMDVPCNHYLTTCCKGTVACQPPVVENAKTFGK 2312

```

A/Accession: A47171
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-3562 <SH>
A:Cross-references: GB:D13542; NID:g391643; PIDN:BA002742.1; PID:g391644
A:Experimental source: stage 22-23 developing limb buds
A:Note: sequence extracted from NCBI backbone (NCBIN:134457)
C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein;
F:166-243/Domain: link protein repeat homology <LNK1>
F:264-345/Domain: link protein repeat homology <LNK2>
F:3258-3289/Domain: EGF homology <EGF1>
F:3296-3327/Domain: EGF homology <EGF>
F:3334-3454/Domain: C-type lectin homology <LCH>
F:3461-3517/Domain: complement factor H repeat homology <FHD>

Query Match          15.3%; Score 171; DB 2; Length 3562;
Best Local Similarity 28.5%; Pred. No. 2.5e-06;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps

QY 17 QPCYKVIYFHDTSRLNFEAKEACRRDGGOLGVSEDEQKLEKFIENLLPSDGF- 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3342 QGOCYK--YF---AHRRTWDARECRLOGAHLTSLISHEQVFNRI-----GHDYQ 3389

QY 76 WIGLRRREKQSNSTACODLYATWDTGSISQFRNWTVDSP-----SCGSEVCVVMYHQPSAP 131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3390 WIGL-----NDKMFERDFRTWDTGSPLOYENWRPNQDFFSAGEDCVVIWHENG-- 3439

QY 132 AGTGGPYMFQWDDNRCKNMKNFKICKYS-----DEKPAVPSRAEGE 172
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3440 -----QWNDVPCNYHLTYTCKKGTVACQGPVWENAKTFEGK 3475

RESULT 10
T42710
mannose receptor, macrophage - mouse
N1 alternate names: lambda lectin; phospholipase A2 receptor
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #start chgno 00-Jan-2000

```

```

A:Reference number: A91327; MUID:85027710; PMID:6489519
A:Accession: E29164
A:Molecule type: protein
C:Residues: 1230-1249 <PE2>
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology
C:Keywords: glycoprotein
F:8-28/Domain: link protein repeat homology (fragment) <LNK1>
F:29-58/Domain: link protein repeat homology (fragment) <LNK2>
F:80-146/Domain: link protein repeat homology (fragments) <LNK3>
F:167-248/Domain: link protein repeat homology <LNK4>
F:1130-1250/Domain: C-type lectin homology <LCH>
F:11257-1313/Domain: complement factor H repeat homology <FHD>

Query Match      14.2%; Score 158.5; DB 2; Length 1340;
Best Local Similarity 26.8%; Pred. No. 1e-05;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

Qy 9 QPVCRG---TORPYKVIYFHTSRRLNFEPAKEACRRDGGQGVLSIESDEQKLEKFI 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1127 QKLCBEGWTKFGCHYCR--HFPPD---RATWDAESQCRKQQSHLSLSTVPEEQ-----EFV 1177

Qy 66 ENLLPSDGF-WIGLRRREEKQSNSTACODLYAWTDGSIQFRNYYVDEP-----SCGSEV 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1178 NN---NAPDYQWIGL-----NDKTIEGDFRWSGHSGLQFENWRPNQDNPFATGDC 1226

Qy 121 CVVMYHQPSAPAGIPGYMFQWNDRCNMKNFICKYS-----DEKPAVPSREAEGETE 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1227 VVMITHEKG-----ENWDVPCNVQLPFTCKKGTGTVACCEPPVVEHARIFGKKD 1274

```

```

RESULT 12
T42630
aggreCAN - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42630
R:Hering, T.M.; Kollar, J.; Huynh, T.D.
submitted to the EMBL Data Library, September 1996
A:Description: Complete coding sequence of bovine aggreCAN: comparative structural
A:Reference number: 222182
A:Accession: T42630
A:Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-2327 <R>
A:Cross-references: EMBL:U76615; NID:g1730259; PID:g1730260; PIDN:AAB38524.1
A:Experimental source: articular chondrocytes
C:Superfamily: aggreCAN; C-type lectin homology; complement factor H repeat homology
C:Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glyc
Query Match 14.2%; Score 158.5; DB 2; Length 2327;
Best Local Similarity 26.8%; Pred. No. 1.9e-05;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;
QY 9 QPVCRGG---TQRPCYKVIYFHDTSRRILNFEAKRDRDGGQLVSIKLEKFI 65
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2114 QKLCEGWTKFGHCYR--HFPD---RATWVDAESQCRKQSHLSIVTPEQ----EFV 2164
| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 ENLLPSDGF--WGLRRREEKQSNACQDLYAWTDGSIQFRNMYVDPE-----SCGSEV 120
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2165 NN---NAQDQWIGL-----NDKTIEGDFRWSGDHSLQFENWRPNQDFNFATGDC 2213
| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 CVVMYHQPSPAGIGGPFMFQWMDRCKNMKNFICKYS----DEKPAVPSREAEGEETE 175
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2214 VVMIIHEKG-----EMWDVPCNYQLPFTCKKGTVACGEPVVEHARIFGOKKD 2261
| | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 13
A39086
aggreCAN precursor, cartilage long splice form [validated] - human
N:Alternate names: chondroitin sulfate proteoglycan 1; large aggregating proteoglyc
N:N/Contains: aggreCAN cartilage short splice form
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

```

```

RESULT 12
T42630
aggreCAN - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42630
R:Hering, T.M.; Kollar, J.; Huynh, T.D.
submitted to the EMBL Data Library, September 1996
A:Description: Complete coding sequence of bovine aggreCAN: comparative structural
A:Reference number: Z22182
A:Accession: T42630
A>Status: preliminary; translated from GB/EMBL/DDBBJ
A:Molecule type: mRNA
A:Residues: 1-2327 <>ER>
A:Cross-references: EMBL:U76615; NID:g1730259; PID:g1730260; PIDN:AAB38524.1
A:Experimental source: articular chondrocytes
C:Superfamily: aggreCAN; C-type lectin homology; complement factor H repeat homolog
C:Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glyco
Query Match          14.2%; Score 158.5; DB 2; Length 2327;
Best Local Similarity 25.8%; Pred. No. 1.9e-05;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

QY      9 QPVCRRGG---TQRPCYKVIYFHDTSRRRLNFEEAKACRRDGGQLVSIEDSEOKLTKEFI 65
       |||   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db    2114 QKLCEGWTKFGHCYR--HFPD---RATWVDAESQCRRQQSHLSIVTPEEQ----EFV 2164

QY      66 ENLLPSGDGF-WIGLRREEKQSNTACODLYAWTDGSGISQFRNYWVDPP-----SCGSEV 120
       |||   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db    2165 NN---NAQDVQMIGL-----NDKTIEGDFRWSDGHSLQFNWRPNOPDNFFATGDC 2213

QY     121 CVVMYHQPSPAGITGGPYMFOWNDRCNMKNFNICKYS----DEKPAPVSREAEGEETE 175
       ||||   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db    2214 VVMIIHEKG-----EWNDVPCNYQLPTCKKGTVACGEPPVVEHARIFGOKKD 2261

RESULT 13
A39086
aggreCAN precursor, cartilage long splice form [validated] - human
N:Alternate names: chondroitin sulfate proteoglycan 1; large aggregating proteoglyc
N:N/Contains: aggreCAN cartilage short splice form
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

```


A:Cross-references: GB:J05062; NID:g181167; PIDN:AAA35726.1; PID:g181168

C:Genetics:

A:Gene: GDB:AGCL1; CSPG1; CSPGCP; MSK16

A:Cross-references: GDB:127479; OMIM:155760

A:Map position: 15q26-15q26

C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology

C:Keywords: alternative splicing; cartilage; chondroitin sulfate proteoglycan; extracellular matrix

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-2415/Product: aggrecan cartilage long splice form #status predicted <MATL>

F:20-2162,2201-2415/Product: aggrecan cartilage short splice form #status predicted <MATL>

F:126,239,333,387,434,602,657,737,1898/Binding site: carbohydrate (Asn) (covalent)

F:317,376/Binding site: keratan sulfate (Thr) (covalent) #status predicted

Query Match 14.2%; Score 158.5; DB 1; Length 2415;
Best Local Similarity 29.1%; Pred. No. 2e-05;
Matches 52; Conservative 22; Mismatches 62; Indels 43; Gaps 11;

QY 9 QPVCRRG--TQPCYKVIYFHDTSRRLLNFEAKACRRDGGQLVSIKLEKFI 65
Db 2202 QEVCEEGNKKYGGHCYR--HFDP-----RETWDAERRCRQSHLSIVTPEQ-----EFV 2252

QY 66 ENLLPSDGF-WIGLRRREKQSNSTACQDLVATDGTSGISQFRNVTYVDSPS---CGSEVC 121
Db 2253 NN---NAQDYQWIGL-----NDRTEGDFRNSDGHFPMQFNWRPNQDNFFAAGEDC 2301

QY 122 VVM-YHQPAPAGIGGYPYFQWNRDRCNMKNFNICKYS----DERPAVPSREAEGETE 175
Db 2302 VVMTHWKG-----EWNDVPCNVYHLPFTCKKGTVACGEPVVEHARTFGQKKD 2349

RESULT 14

B42755

E-selectin precursor - mouse

N:Alternate names: endothelial leukocyte adhesion molecule 1 (ELAM-1)

C:Species: Mus musculus (house mouse)

C:Date: 04-Mar-1993 #sequence_revision 31-Dec-2000 #text_change 02-Aug-2002

C:Accession: S23174; B42755

R:Becker-Andre, M.; van Huijsduijn, R.H.; Losberger, C.; Whelan, J.; Delamar, J.

A:Title: Murine endothelial leukocyte-adhesion molecule 1 is a close structural and functional homologue of human endothelial leukocyte-adhesion molecule 1

A:Reference number: S23174; MUID:92283265; PMID:1375914

A:Accession: S23174

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-612 <BEC>

A:Cross-references: GB:M8778; NID:g193014; PIDN:AAA37547.1; PID:g193015

R:Weller, A.; Isenmann, S.; Vestweber, D.

J. Biol. Chem. 267, 15176-15183, 1992

A:Title: Cloning of the mouse endothelial selectins. Expression of both E- and P-selectin in endothelial cells

A:Reference number: A42755; MUID:92340571; PMID:1378846

A:Accession: B42755

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: "MKATAGV", 1-389, 391-612 <WEL>

A:Cross-references: GB:M87862; NID:g193107

A:Experimental source: endothelial cells

A>Note: sequence extracted from NCBI backbone (NCBI:109470)

A>Note: it is uncertain whether the initiator is Met-1 or the AUG codon preceding the initiator

C:Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology

C:Keywords: glycoprotein; transmembrane protein

ИЗДАТЕЛЬСТВО «НАУКА»

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:00:15 ; Search time 13 Seconds
(without alignments)
657.240 Million cell updates/sec

Title: US-09-887-855-5

Perfect score: 1115

Sequence: 1 ATGRLSGQVCRGGTORPC.....BEDAKTKFKSREAAALNAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	574	51.5	273	1	Q9cxm0 mus musculus
2	561.5	50.4	273	1	CHOD_MOUSE
3	182	16.3	1456	1	MANR_HUMAN
4	177	15.9	1268	1	PGCN_MOUSE
5	174.5	15.7	3381	1	PGCV_BOVIN
6	174	15.6	1257	1	PGCN_RAT
7	174	15.6	2738	1	PGCV_RAT
8	174	15.6	3358	1	PGCV_MOUSE
9	174	15.6	3396	1	PGCV_HUMAN
10	171	15.3	3562	1	PGCV_CHICK
11	165	14.8	643	1	CD93_RAT
12	158.5	14.2	2364	1	PGCA_BOVIN
13	158.5	14.2	2415	1	PGCA_HUMAN
14	155.5	13.9	2333	1	PGCA_CANFA
15	154.5	13.9	612	1	LEM2_MOUSE
16	153.5	13.8	644	1	CD93_MOUSE
17	153.5	13.8	2124	1	PGCA_RAT
18	152	13.6	912	1	PGCB_BOVIN
19	151.5	13.6	652	1	CD93_HUMAN
20	151	13.5	197	1	CLF1_HUMAN
21	149	13.4	321	1	FCE2_HUMAN
22	148.5	13.3	2132	1	PGCA_MOUSE
23	147	13.2	883	1	PGCB_MOUSE
24	146	13.1	2109	1	PGCA_CHICK
25	145	13.0	158	1	LECG_TRIST
26	145	13.0	883	1	PGCB_RAT
27	144.5	13.0	173	1	LEC2_MGRO
28	144.5	13.0	372	1	LEM1_RAT
29	143.5	12.9	372	1	LEM1_MOUSE
30	141.5	12.7	331	1	FCE2_MOUSE
31	141.5	12.7	549	1	LEM2_RAT
32	140.5	12.6	162	1	LEC3_MGRO
33	140	12.6	370	1	LEM1_BOVIN
					P98131 bos taurus

34	139	12.5	248	1	PSPA_HUMAN
35	139	12.5	283	1	LECA_SARPE
36	138.5	12.4	152	1	IXA_TRIFL
37	138.5	12.4	202	1	TETN_MOUSE
38	137.5	12.3	372	1	LEM1_MACMU
39	137.5	12.3	372	1	LEM1_PAPHA
40	136	12.2	175	1	LITH_BOVIN
41	135	12.1	165	1	LIT1_MOUSE
42	134.5	12.1	132	1	ACAL_ANGAN
43	134.5	12.1	372	1	LEM1_PONPY
44	134.5	12.1	485	1	LEM2_BOVIN
45	134	12.0	166	1	TETN_CARSP

ALIGNMENTS

RESULT 1
CHOD_MOUSE
ID CHOD_MOUSE STANDARD; PRT; 273 AA.
AC Q9CXM0; Q8VI31;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chondrolectin precursor (Transmembrane protein MT75).
GN CHODL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Weig L., Smits P., Hubner R., Wouters J., Merregaert J.;
RT "Mt75, a low expressed c-type lectin gene involving in
RT chondrogenesis";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
Functional annotation of a full-length mouse cDNA collection;
Nature 409:685-690(2001).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF311699; AAL50354.1; -.


```

Db 23 RVSGQKVCFAFKHPCYKMAFYHELSSRVSFQEARLACESGGVLLSLENEAEQKLIES 82
OY 64 FTENLLP-----SDGDFWIGLRRRREKQSNSTACODLYAWTDGSIQPRNHYVDEPSCGS 118
Db 83 MLQNTKCTGSLGDFWIGLWRNGDGT-SGACPDLYQWSDGNSQYRNWYTDPEPSCGS 141
OY 119 EYCVVYHQPSAPAGIGGYFQWDDRCNKNFNICKYSDF-KPAVP 165
Db 142 EKVVMYHPTANPLGGLGYLQWDDRCNKNHNYICKYPEINTAP 189

RESULT 3
ID MANR_HUMAN STANDARD; PRT; 1456 AA.
AC P22897;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Macrophage mannose receptor precursor (MMR) (CD206 antigen).
GN MRC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=90324192; PubMed=2373685;
RA Taylor M.E., Conary J.T., Lennartz M.R., Stahl P.D., Drickamer K.;
RT "Primary structure of the mannose receptor contains multiple motifs
RT resembling carbohydrate-recognition domains.";
RL J. Biol. Chem. 265:12156-12162(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052405; PubMed=1294118;
RA Kim S.-J., Ruiz N., Bezouska K., Drickamer K.;
RT "Organization of the gene encoding the human macrophage mannose
RT receptor (MRC1).";
RL Genomics 14:721-727(1992).
RN [3]
RP STUDIES ON THE BINDING OF INDIVIDUAL LECTIN DOMAINS.
RX MEDLINE=92112893; PubMed=1730714;
RA Taylor M.E., Bezouska K., Drickamer K.;
RT "Contribution to ligand binding by multiple carbohydrate-recognition
RT domains in the macrophage mannose receptor.";
RL J. Biol. Chem. 267:1719-1726(1992).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 642-788.
RX MEDLINE=20347275; PubMed=10779515;
RA Feinberg H., Park-Snyder S., Kolatkar A.R., Helse C.T., Taylor M.E.,
RA Weis W.I.;
RT "Structure of a C-type carbohydrate recognition domain from the
RT macrophage mannose receptor.";
RL J. Biol. Chem. 275:21539-21548(2000).
CC -1- FUNCTION: MEDIATES THE ENDOCYTOSIS OF GLYCOPROTEINS BY
CC MACROPHAGES, IN SEVERAL RECOGNITION AND UPTAKE PROCESSES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: CRDS 1-3 HAVE AT MOST VERY WEAK AFFINITY FOR
CC CARBOHYDRATE. CRD 4 SHOWS THE HIGHEST AFFINITY BINDING AND HAS
CC MULTISPECIFICITY FOR A VARIETY OF MONOSACCHARIDES. AT LEAST 3 CRDS
CC (4, 5, AND 7) ARE REQUIRED FOR HIGH AFFINITY BINDING AND
CC ENDOCYTOSIS OF MULTIVALENT GLYCOCONJUGATES.
CC -1- SIMILARITY: CONTAINS 8 C-TYPE LECTIN FAMILY DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
CC -1- DATABASE: NAME=PROW; NOTE=PROW 2:85-89(2001);
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1644341535-g.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05550; AAA59868.1; -
DR EMBL; M93221; AAA60389.1; -
DR EMBL; M93192; AAA60389.1; JOINED.
DR EMBL; M93193; AAA60389.1; JOINED.
DR EMBL; M93194; AAA60389.1; JOINED.
DR EMBL; M93195; AAA60389.1; JOINED.
DR EMBL; M93196; AAA60389.1; JOINED.
DR EMBL; M93197; AAA60389.1; JOINED.
DR EMBL; M93198; AAA60389.1; JOINED.
DR EMBL; M93199; AAA60389.1; JOINED.
DR EMBL; M93200; AAA60389.1; JOINED.
DR EMBL; M93201; AAA60389.1; JOINED.
DR EMBL; M93202; AAA60389.1; JOINED.
DR EMBL; M93203; AAA60389.1; JOINED.
DR EMBL; M93204; AAA60389.1; JOINED.
DR EMBL; M93205; AAA60389.1; JOINED.
DR EMBL; M93206; AAA60389.1; JOINED.
DR EMBL; M93207; AAA60389.1; JOINED.
DR EMBL; M93208; AAA60389.1; JOINED.
DR EMBL; M93209; AAA60389.1; JOINED.
DR EMBL; M93210; AAA60389.1; JOINED.
DR EMBL; M93211; AAA60389.1; JOINED.
DR EMBL; M93212; AAA60389.1; JOINED.
DR EMBL; M93213; AAA60389.1; JOINED.
DR EMBL; M93214; AAA60389.1; JOINED.
DR EMBL; M93215; AAA60389.1; JOINED.
DR EMBL; M93216; AAA60389.1; JOINED.
DR EMBL; M93217; AAA60389.1; JOINED.
DR EMBL; M93218; AAA60389.1; JOINED.
DR EMBL; M93219; AAA60389.1; JOINED.
DR EMBL; M93220; AAA60389.1; JOINED.
DR PIR; A36563; A36563.
DR PDB; 1EGG; 30-AUG-00.
DR PDB; 1EGI; 30-AUG-00.
DR Genew; HGNC:7228; MRC1.
DR MIM; 153618; -
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR PRINTS; PR00013; FNYPEII.
DR ProDom; PD000995; FN_Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 6.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 8.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
DR Receptor; Signal; Calcium-binding; Transmembrane; Repeat;
DR Glycoprotein; Endocytosis; Macrophage; Lectin; Antigen; 3D-structure.
FT SIGNAL 1 18
FT CHAIN 19 1456
FT DOMAIN 19 1383
FT TRANSMEM 1384 1411
FT DOMAIN 1412 1456
FT DOMAIN 22 142
FT DOMAIN 157 212
FT DOMAIN 216 344
FT DOMAIN 360 490
FT DOMAIN 502 629
FT DOMAIN 644 781
FT DOMAIN 805 926
FT DOMAIN 943 1083
FT DOMAIN 1100 1216
FT DOMAIN 1228 1359
FT DISULFID 646 659
FT DISULFID 680 777
FT DISULFID 753 769

```

```

FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 930 930 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1160 1160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1205 1205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1456 AA; 166011 MW; 2645AF3C576A5E3 CRC64;

Query Match 16.3%; Score 182; DB 1; Length 1456;
Best Local Similarity 25.5%; Pred. No. 2.8e-08;
Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;

Qy 21 YKVIYFHTDTRRLNPFEEAKACRRDGGQLVSISEDEQKLEKFIENLLPSDGDGFWIGLR 80
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 807 YKDYVFSKEKMDNARAFCKRNFGLVSIQSESEKKFLWKYV-NRNDQAQSAVFICLL 865

Qy 81 RREKOSNSTACDLYAWTDGSIQFRNRYVDEPSCGS--EVCVVMYHQPAPAGIGGPY 138
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 866 ISLKK-----FAMDGSKVDYVSWATGEFNANEDENCVTMY-----SNSGF----- 908

Qy 139 MFOVNDRCNNKFNICKYSDK-----PAVPSREAEGEETELTPVLPEETQE----- 187
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 909 ---NNDINGYPNAFICORHNSINATVMP-----TMPSPSGCKEGNWFYSN 954

Qy 188 -----EAKTFKESRAAL 202
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 955 KCFKIFGMEERKNWQEARACI 978

RESULT 4
PGCN_MOUSE STANDARD; PRT; 1268 AA.
AC P35066;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurocan core protein precursor.
GN CSPG3 OR NCAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=Brain;
RX MEDLINE=96039250; PubMed=7490074;
RA Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
RA Faessler R.;
RT "Structure and chromosomal localization of the mouse neurocan gene.";
RL Genomics 28:405-410(1995).
CC -1- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
CC DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM
CC AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
CC ACID.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X8459216.1; -

```

```

DR HSSP; P00740; LEDM.
DR MGD; MGI:104694; Cspg3.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; Lectin_C; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_Ca; 1.
DR SMART; SM00001; EGF_Like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_Ca; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1268 NEUROCAN CORE PROTEIN.
FT DOMAIN 51 146 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 176 253 LINK 1.
FT DOMAIN 274 355 LINK 2.
FT DOMAIN 960 996 EGF-LIKE 1.
FT DOMAIN 998 1034 EGF-LIKE 2.
FT DOMAIN 1036 1165 C-TYPE LECTIN.
FT DOMAIN 1166 1224 SUSHI.
FT DISULFID 58 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 354 BY SIMILARITY.
FT DISULFID 303 324 BY SIMILARITY.
FT DISULFID 964 975 BY SIMILARITY.
FT DISULFID 969 984 BY SIMILARITY.
FT DISULFID 986 995 BY SIMILARITY.
FT DISULFID 1040 1051 BY SIMILARITY.
FT DISULFID 1068 1160 BY SIMILARITY.
FT DISULFID 1136 1152 BY SIMILARITY.
FT DISULFID 1167 1210 BY SIMILARITY.
FT DISULFID 1196 1223 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1175 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1268 AA; 137200 MW; 3014E8E202A2FAEC CRC64;

```

Query Match 15.9%; Score 177; DB 1; Length 1268;
Best Local Similarity 31.5%; Pred. No. 6.6e-08;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;

```

Qy 17 QRCYKVIYFHTDTRRLNPFEEAKACRRDGGQLVSISEDEQKLEKFIENLLPSDGDGFW 76
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1048 QGHYR--YF---AHRRAWEDAEDCRRRAGHLTSVHSPEEHKFINSF-----GHENSW 1096

Qy 77 IGLRRREKOSNSTACDLYAWTDGSIQFRNRYVDEPS---CGSEVCVVMYHQPAPAG 133
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1097 IGLNDRIVERD-----FOWTDNTGLQYENWREKQPDNFFAGGEDCVVVAHESG--- 1145

```

```

QY 134 IGGPYMFQWQNDRCNNKNNFICK 156
DB 1146 -----RWNDVPCYNLPYVCK 1161

RESULT 5
PCGV_BOVIN
ID PGCV_BOVIN STANDARD; PRT: 3381 AA.
AC P81282; O77609; O77610; O77611; O77612;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Versican core protein precursor (large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (CHAP).
GN CPG2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS V0; V1; V2 AND V3).
RC TISSUE=Forebrain;
RX MEDLINE=98288320; PubMed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764(1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348.
RX TISSUE=Spinal cord;
RX MEDLINE=92062692; PubMed=1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid";
RL Biochim. Biophys. Acta 1075:248-258(1991).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2
CC and V3; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF060456; AAC24358.1; -
CC EMBL; AF060457; AAC24359.1; -
CC EMBL; AF060458; AAC24360.1; -
CC EMBL; AF060459; AAC24361.1; -
CC HSSP; P01132; IEPG.
CC InterPro; IPR000152; Asx_hydroxyl.

```

```

DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00047; Ig_1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 1.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF_2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00409; IG_1.
DR SMART; SM00410; IG_Like; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00411; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3381 IC-LIKE V-TYPE DOMAIN.
FT DOMAIN 37 138 LINK 1.
FT DOMAIN 168 245 LINK 2.
FT DOMAIN 266 347 LINK 2.
FT DOMAIN 349 1336 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT DOMAIN 1337 3074 DOMAIN).
FT DOMAIN 3074 3110 GAG-BETA.
FT DOMAIN 3112 3148 EGF-LIKE 1.
FT DOMAIN 3161 3275 EGF-LIKE 2, CALCIUM-BINDING.
FT DOMAIN 3280 3338 C-TYPE LECTIN.
FT DISULFID 44 131 SUSHI.
FT DISULFID 173 244 BY SIMILARITY.
FT DISULFID 197 218 BY SIMILARITY.
FT DISULFID 271 346 BY SIMILARITY.
FT DISULFID 295 316 BY SIMILARITY.
FT DISULFID 3078 3098 BY SIMILARITY.
FT DISULFID 3083 3098 BY SIMILARITY.
FT DISULFID 3100 3109 BY SIMILARITY.
FT DISULFID 3116 3127 BY SIMILARITY.
FT DISULFID 3121 3136 BY SIMILARITY.
FT DISULFID 3138 3147 BY SIMILARITY.
FT DISULFID 3154 3165 BY SIMILARITY.
FT DISULFID 3182 3274 BY SIMILARITY.
FT DISULFID 3250 3266 BY SIMILARITY.
FT DISULFID 3281 3324 BY SIMILARITY.
FT DISULFID 3310 3337 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1017 1017 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1333 1333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1463 1463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1653 1653 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1974 1974 N-LINKED (GLCNAC. . .) (POTENTIAL).

```


Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
 "Proteoglycan expression in the normal rat kidney.";
 Nephron 77:461-470(1997).
 [4]
 SEQUENCE OF 2535-2738 FROM N.A.
 STRAIN-Sprague-dawley; TISSUE-Lung;
 Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
 "Molecular cloning and characterization of two developmentally
 regulated genes in rat lung.";
 Submited (SP-2000) to the EMBL/GenBank/DBJ databases.
 -!- FUNCTION: May play a role in intercellular signaling and in
 connecting cells with the extracellular matrix. May take part in
 the regulation of cell motility, growth and differentiation. Binds
 hyaluronic acid.
 -!- SUBCELLULAR LOCATION: secreted; extracellular matrix.
 -!- ALTERNATIVE PRODUCTS: At least 3 isoforms; V0 (shown here), V3 and
 Vint; are produced by alternative splicing. In the papillary area,
 -!- TISSUE SPECIFICITY: In kidney is expressed in the papillary area,
 but not in glomeruli.
 -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
 (by similarity).
 -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 -!- SIMILARITY: CONTAINS 1 SCR) DOMAIN.
 -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF062402; AAC40166.1; -
 EMBL; U75306; AAB51125.1; -
 EMBL; AF084544; AAD48544.1; -
 EMBL; AF072892; AAC26116.1; -
 EMBL; AY007691; AAG16631.1; -
 HSP; P01132; IEPG.
 InterPro: IPR000152; Asx_hydroxyl.
 InterPro: IPR000561; EGF-like.
 InterPro: IPR000742; EGF_2.
 InterPro: IPR001881; EGF_Ca.
 InterPro: IPR003595; Ig.
 InterPro: IPR003600; Ig-like.
 InterPro: IPR001304; Lectin_C.
 InterPro: IPR000538; Link.
 InterPro: IPR000436; Sushi_SCR_CCP.
 PRINTS; PR01265; LINKMODULE.
 SMART; SM00032; CCP; 1.
 SMART; SM00034; CLECT; 1.
 SMART; SM00181; EGF; 2.
 SMART; SM00179; EGF_Ca; 2.
 SMART; SM00409; IG; 1.
 SMART; SM00410; IG-like; 1.
 SMART; SM00445; LINK; 2.
 PROSITE; PS00010; ASX_HYDROXYL; 1.
 PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
 PROSITE; PS00022; EGF_1; 2.
 PROSITE; PS01186; EGF_2; 1.
 PROSITE; PS01187; EGF_Ca; 1.
 PROSITE; PS01241; LINK; 2.
 Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 Hyaluronic acid; Alternative splicing
 SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2738
 FT NON_CONS 348 349
 FT DOMAIN 37 137
 IG-LIKE V-TYPE DOMAIN.

(Chondroitin sulfate proteoglycan core protein 2) (PG-M).

CSPE2.

MS musculus (Mouse).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

[1]

SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).

STRAIN=C57BL/6, and Swiss Webster; TISSUE=Brain;

MEDLINE=951122551; PubMed=782336;

Ita K., Shinomura T., Zako M., Ujita M., Kinata K.;

"Multiple forms of mouse PG-M, a large chondroitin sulfate proteoglycan generated by alternative splicing.";

J. Biol. Chem. 270:958-965(1995).

[2]

SEQUENCE FROM N.A. (ISOFORM V3).

STRAIN=C57BL/6;

MEDLINE=95181355; PubMed=7876137;

Zako M., Shinomura T., Ujita M., Ito K., Kinata K.;

"Expression of PG-M(V3), an alternatively spliced form of PG-M without a chondroitin sulfate attachment in region in mouse and human tissues.";

J. Biol. Chem. 270:3914-3918(1995).

[3]

SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).

STRAIN=C57BL/6J; TISSUE=Skin;

MEDLINE=21085660; PubMed=11217851;

Kawai J., Shinagawa A., Shibata K., Yoshino M., Itochi M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann C., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;

"Functional annotation of a full-length mouse cDNA collection.";

Nature 409:585-690(2001).

-1- FUNCTION: May play a role in intercellular signaling and in connecting cells with the extracellular matrix. May take part in the regulation of cell motility, growth and differentiation. Binds hyaluronic acid.

-1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.

-1- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2 and V3; are produced by alternative splicing.

-1- TISSUE SPECIFICITY: V2 is found only in brain.

-1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.

-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

-1- SIMILARITY: CONTAINS 2 LINK DOMAINS.

-1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.

-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

-1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.

-1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcements> or send an email to license@sib-sib.ch).

EMBL; D16263; BAA03796.1; ;

DR EMBL; D28599; ; NOT ANNOTATED CDS.

```

DR EMBL; D32040; BAA06802.1; .
DR EMBL; AK014525; BAB29411.1; .
DR HSP; P01132; 1EPG.
DR MGD; MGI:102889; Cspg2.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBL00D.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CLEC; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT CHAIN 1 20
FT DOMAIN 21 3358
FT DOMAIN 37 137
FT DOMAIN 167 244
FT DOMAIN 265 346
FT DOMAIN 348 1308
FT DOMAIN 1309 3052
FT DOMAIN 3052 3088
FT DOMAIN 3090 3126
FT DOMAIN 3139 3253
FT DOMAIN 3258 3316
FT DISULFID 44 130
FT DISULFID 172 243
FT DISULFID 196 217
FT DISULFID 270 333
FT DISULFID 294 315
FT DISULFID 3056 3067
FT DISULFID 3061 3076
FT DISULFID 3078 3087
FT DISULFID 3094 3105
FT DISULFID 3099 3114
FT DISULFID 3116 3125
FT DISULFID 3132 3143
FT DISULFID 3160 3252
FT DISULFID 3228 3244
FT DISULFID 3259 3302
FT DISULFID 3288 3315
FT CARBOHYD 57 57
FT CARBOHYD 330 330
FT CARBOHYD 351 351
FT CARBOHYD 441 441
FT CARBOHYD 807 807
FT CARBOHYD 914 914
FT CARBOHYD 951 951
FT CARBOHYD 1305 1305
DR EMBL; D32040; BAA06802.1; .
DR EMBL; AK014525; BAB29411.1; .
DR HSP; P01132; 1EPG.
DR MGD; MGI:102889; Cspg2.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBL00D.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CLEC; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT CHAIN 1 20
FT DOMAIN 21 3358
FT DOMAIN 37 137
FT DOMAIN 167 244
FT DOMAIN 265 346
FT DOMAIN 348 1308
FT DOMAIN 1309 3052
FT DOMAIN 3052 3088
FT DOMAIN 3090 3126
FT DOMAIN 3139 3253
FT DOMAIN 3258 3316
FT DISULFID 44 130
FT DISULFID 172 243
FT DISULFID 196 217
FT DISULFID 270 333
FT DISULFID 294 315
FT DISULFID 3056 3067
FT DISULFID 3061 3076
FT DISULFID 3078 3087
FT DISULFID 3094 3105
FT DISULFID 3099 3114
FT DISULFID 3116 3125
FT DISULFID 3132 3143
FT DISULFID 3160 3252
FT DISULFID 3228 3244
FT DISULFID 3259 3302
FT DISULFID 3288 3315
FT CARBOHYD 57 57
FT CARBOHYD 330 330
FT CARBOHYD 351 351
FT CARBOHYD 441 441
FT CARBOHYD 807 807
FT CARBOHYD 914 914
FT CARBOHYD 951 951
FT CARBOHYD 1305 1305

```

```

FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1679 1679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2054 2054 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2244 2244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2362 2362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2627 2627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3030 3030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3332 3332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3342 3342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 348 348 P -> R (IN ISOFORM V1 AND ISOFORM V3).
FT VARSPLIC 349 1308 MISSING (IN ISOFORM V1).
FT VARSPLIC 1309 3052 MISSING (IN ISOFORM V2).
FT VARSPLIC 349 3052 MISSING (IN ISOFORM V3).
FT CONFLICT 126 126 A -> G (IN REF. 3).
FT CONFLICT 348 348 MISSING (IN REF. 3).
FT CONFLICT 1658 1658 I -> T (IN REF. 3).
FT CONFLICT 1674 1680 TVMNSNS -> QFGIQA (IN REF. 3).
SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;

Query Match 15.6%; Score 174; DB 1; Length 3358;
Best Local Similarity 28.5%; Pred. No. 3.8e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

QY 17 QRPCKYVYFHDTSRLNFEFAKACRDGGQLVSVESIEDEQKLIKFIENLLPSDGF- 75
DB 3140 QGQCYK-YF---AHRRTWDAARECRLOGAHLTSLSHEEQMFVNRV-----GHDYQ 3187
QY 76 WIGLRREKQSNSTACQDLYAWTQGISQFRNRYVDEP-----SCSEVCVVMYHPSAP 131
DB 3188 WIGL-----NDKMFDEHDFRMTDGSALQYENRPNQDFFSAGEDCVIWHENG-- 3237
QY 132 AGIGGPFYFQWDCRNKNFKYKYS---DEKPAVPSREARGE 172
DB 3238 -----QWNDVPCNYHLTYTCKKGTVACGQPPVVENAKTFGR 3273

RESULT 9
PCGV_HUMAN
ID PGCV_HUMAN STANDARD; PRT: 3396 AA.
AC P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia)
DE hyaluronate-binding protein) (GHA).
GN CSPE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM V0).
RX MEDLINE=95105188; PubMed=7528742;
RA Naso M.F.; Zimmermann D.R.; Iozzo R.V.;
RT "Characterization of the complete genomic structure of the human
RL J. Biol. Chem. 269:32999-33008(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM V1).
RC TISSUE=Placenta;
RX MEDLINE=90059882; PubMed=2583089;
RA Zimmermann D.R.; Ruoslahti E.;
RT "Multiple domains of the large fibroblast proteoglycan, versican.";
EMBO J. 8:2975-2981(1989).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM V2).
RX MEDLINE=95105187; PubMed=7806529;
RA Dours-Zimmermann M.F.; Zimmermann D.R.;
RT "A novel glycosaminoglycan attachment domain identified in two
RL J. Biol. Chem. 269:32992-32998(1994).

```


FT DISULFID 3265 3281 BY SIMILARITY.
 FT DISULFID 3296 3339 BY SIMILARITY.
 FT DISULFID 3325 3352 BY SIMILARITY.
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 809 809 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1442 1442 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1468 1468 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1663 1663 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2179 2179 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2280 2280 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2360 2360 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 Query Match 15.6%; Score 174; DB 1; Length 3396;
 Best Local Similarity 28.5%; Pred. No. 3.9e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 17 QRCVKYVYFHTSRRLNPEEAKACRRDGGOLVSIESDEQKLEKFIENLLPSDGP- 75
 DB 3177 QGCYK--YF---AHRRTWDAARECRQLQGAHLTSLSHQMFVNRV-----GHDYQ 3224
 QY 76 WGLRRREKQSNSTACQDLYAWTQSGISQFNWYVDEP-----SCGSEVCVWYHQPSPAP 131
 DB 3225 WGL-----NDKMFHDFRWTQSTLOYENWRPNPDSFSSAGEDCVIWHENG-- 3274
 QY 132 AGIGPYMFQWDDRCNNKNNFICKYS-----DEKPAVPSREAGE 172
 DB 3275 -----QNDVFCNVHLTYTCKGTVACGPPVWENAKTFGK 3310
 RESULT 10
 ID PGCV_CHICK STANDARD; PRT: 3562 AA.
 AC Q90953; Q90945;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Versican core protein precursor (large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
 GN CPBG2.
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 CC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Limb bud;
 RA MEDLINE=93300846; PubMed=8314802;
 RA Shinomura T., Nishida Y., Ito K., Kimata K.;
 RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
 RT expressed during chondrogenesis in chick limb buds. Alternative
 RT spliced multiforms of PG-M and their relationships to versican.";
 RL J. Biol. Chem. 268:14461-14469(1993).
 CC -1- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms; V0 (shown here) and
 CC V1; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Prechondrogenic condensation area of
 CC developing limb buds.
 CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
 CC (by similarity).
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL: X60226; CAA42787.1; -
 EMBL: D13542; BAA02742.1; -
 HSP: P00740; IEDM.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR000538; Link.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00008; EGF; 2.
 DR Pfam: PF00047; Ig; 1.
 DR Pfam: PF00059; lectin_c; 1.
 DR Pfam: PF00084; sushi; 1.
 DR Pfam: PF00193; Xlink; 2.
 DR ProDom: PD000918; Link; 2.
 DR SMART: SM00032; CCP; 1.
 DR SMART: SM00034; CLECT; 1.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00001; EGF_like; 1.
 DR SMART: SM00409; IG; 1.
 DR SMART: SM00445; LINK; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS01241; LINK; 2.
 DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE: PS00041; C-TYPE_LECTIN_2; 1.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 26
 FT CHAIN 27 3562 VERSICAN CORE PROTEIN.
 FT DOMAIN 37 136 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 166 243 LINK 1.
 FT DOMAIN 264 345 LINK 2.
 FT DOMAIN 3254 3290 EGF-LIKE 1.
 FT DOMAIN 3292 3328 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 3341 3455 C-TYPE LECTIN.
 FT DOMAIN 3460 3518 SUSHI.
 FT DISULFID 44 129 BY SIMILARITY.
 FT DISULFID 171 242 BY SIMILARITY.
 FT DISULFID 195 216 BY SIMILARITY.
 FT DISULFID 269 344 BY SIMILARITY.
 FT DISULFID 293 314 BY SIMILARITY.
 FT DISULFID 3258 3269 BY SIMILARITY.
 FT DISULFID 3263 3278 BY SIMILARITY.
 FT DISULFID 3280 3289 BY SIMILARITY.
 FT DISULFID 3296 3307 BY SIMILARITY.
 FT DISULFID 3301 3316 BY SIMILARITY.
 FT DISULFID 3318 3327 BY SIMILARITY.
 FT DISULFID 3334 3345 BY SIMILARITY.
 FT DISULFID 3362 3454 BY SIMILARITY.
 FT DISULFID 3430 3446 BY SIMILARITY.
 FT DISULFID 3461 3504 BY SIMILARITY.
 FT DISULFID 3490 3517 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).


```

DR DR PROSITE; PS01187; EGF_CA; 1.
DR DR PROSITE; PS00290; IG_MHC; FALSE_NEG.
DR DR PROSITE; PS01241; LINK; 4.
DR DR PROSITE; PS00615; C-TYPE-LECTIN.1; 1.
DR DR PROSITE; PS50041; C-TYPE-LECTIN.2; 1.
KW KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW EGF-like domain; Calcium; Alternative splicing; Repeat;
KW Immunoglobulin domain.
FT FT SIGNAL 1 16
FT FT CHAIN 17 2364
FT FT DOMAIN 44 140
FT FT DOMAIN 170 247
FT FT DOMAIN 268 349
FT FT DOMAIN 504 581
FT FT DOMAIN 602 687
FT FT DOMAIN 774 903
FT FT DOMAIN 1433 2112
FT FT DOMAIN 2113 2149
FT FT DOMAIN 2114 2364
FT FT DOMAIN 2161 2276
FT FT DOMAIN 2280 2338
FT FT DISULFID 51 133
FT FT DISULFID 175 246
FT FT DISULFID 199 220
FT FT DISULFID 273 348
FT FT DISULFID 297 318
FT FT DISULFID 509 580
FT FT DISULFID 533 554
FT FT DISULFID 607 682
FT FT DISULFID 631 652
FT FT DISULFID 2117 2128
FT FT DISULFID 2182 2278
FT FT DISULFID 2250 2266
FT FT DISULFID 2281 2324
FT FT DISULFID 2310 2337
FT FT CARBOHYD 126 126
FT FT CARBOHYD 239 239
FT FT CARBOHYD 333 333
FT FT CARBOHYD 387 387
FT FT CARBOHYD 611 611
FT FT CARBOHYD 667 667
FT FT VARSPLIC 2114 2150
FT FT VARSPLIC 2364 2364
SQ SEQUENCE 2364 AA; 246359 MW; 6FF83763420C3D4C CRC64;

Query Match 14.2%; Score 158.5; DB 1; Length 2364;
Best Local Similarity 26.8%; Pred. No. 6.1e-06;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

QY 9 QVPCRGG--TORPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSYSEDEQKLEKFI 65
DB 2151 QKUCEGWTQFGHCYR--HFPD---RATWVDAESQCKRQOQSHLSIVTPEQ----EFV 2201

QY 66 ENLLPSDGF-WIGLRREKQSNSTACQDIYAWTDGSGISQFRNYYVDEP----SCGSEV 120
DB 2202 NN---NAQDYQWIGL-----NDKTIEGDFRWSGDGSHLQFENRNPQDNFATGDC 2250

QY 121 CVVMYHPSAPAGTGGPYMFQWNDRCNMKNFTCKYS----DEKPAVPSREAPGEETE 175
DB 2251 VVMITHEKG-----EWNDVPCNYQLPFTCKRKTGACGPPVVEHARIFGQK 2298

RESULT 13
PGCA_HUMAN STANDARD; PRT; 2415 AA.
ID PGCA_HUMAN
AC P16112; Q13650;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPC) (Chondroitin sulfate proteoglycan core protein 1).
GN AGC1 OR CSPC1.
OS Homo sapiens (Human).

```

```

OC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN CC NCBI_TaxID=9606;
RP CC SEQUENCE OF 1-2162; 2201-2329 AND 2391-2415 FROM N.A.
RC CC TISSUE=Chondrocytes;
RX CC MEDLINE=91093289; PubMed=1985970;
RA CC Doege K.J., Sasaki M., Kimura T., Yamada Y.;
RT CC "Complete coding sequence and deduced primary structure of the human
RT CC cartilage large aggregating proteoglycan, aggrecan. Human-specific
RL CC repeats, and additional alternatively spliced forms.";
RN CC J. Biol. Chem. 266:894-902(1991).
RP CC [2]
RC CC SEQUENCE OF 1778-2162 AND 2201-2415 FROM N.A.
RX CC TISSUE=Chondrocytes;
RA CC Dudhia J., Hardingham T.E.;
RL CC Submitted (JAN-1990) to the EMBL/GenBank/DBJ databases.
RP CC [3]
RN CC SEQUENCE OF 1936-2415 FROM N.A.
RX CC MEDLINE=89380154; PubMed=2789216;
RA CC Baldwin C.T., Reginato A.M., Prockop D.J.;
RT CC "A new epidermal growth factor-like domain in the human core protein
RT CC for the large cartilage-specific proteoglycan. Evidence for
RL CC alternative splicing of the domain.";
RN CC J. Biol. Chem. 264:15747-15750(1989).
RP CC [4]
RC CC SEQUENCE OF 764-864 FROM N.A.
RX CC TISSUE=Blood;
RA CC Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT CC "Length variation in the keratan sulfate domain of mammalian
RL CC aggrecan.";
RN CC Matrix Biol. 14:323-328(1994).
RP CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
RC CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
RX CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
RA CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION.
RT CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
RN CC similarity).
RP CC -1- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST 3 FORMS OF AGGREGAN
RC CC TRANSCRIPTS GENERATED BY ALTERNATIVE EXON USAGE. THE SHORTEST AND
RX CC ALSO THE MOST PREVALENT FORM, LACKS BOTH THE EGF-LIKE DOMAIN, AND
RA CC SUSHI DOMAIN.
RT CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
RN CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
RP CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
RC CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
RX CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
RA CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
RT CC AND G3.
RN CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
RP CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
RX CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
RA CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
RT CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
RN CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
RP CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
RX CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
RN CC -----
RP CC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RX CC the European Bioinformatics Institute. There are no restrictions on its
RA CC use by non-profit institutions as long as its content is in no way
RT CC modified and this statement is not removed. Usage by and for commercial
RL CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN CC or send an email to licensel@isb-sib.ch).
RP CC -----
RX CC EMBL; M55172; AAA62824.1; -
RA CC EMBL; J05062; AAA35726.1; -
RT CC EMBL; X17406; CAA35463.1; -
RN CC EMBL; S74659; AAC60643.2; -
RP CC PIR; S08042; S08042.
RX CC HSSP; P98066; ITSG.

```


DR DR Genew: HGNC:319; AGCL.
 DR DR MIN; 155760; -
 DR DR InterPro: IPR000561; EGF-like.
 DR DR InterPro: IPR003599; IG.
 DR DR InterPro: IPR003006; IG_MHC.
 DR DR InterPro: IPR001304; Lectin_C.
 DR DR InterPro: IPR000538; Link.
 DR DR InterPro: IPR003324; SCXSG.
 DR DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR DR Pfam: PF00008; EGF; 1.
 DR DR Pfam: PF00047; Ig; 1.
 DR DR Pfam: PF00059; lectin_c; 1.
 DR DR Pfam: PF00084; sushi; 1.
 DR DR Pfam: PF00193; Link; 4.
 DR DR Pfam: PF02339; SCXSG; 71.
 DR DR ProDom: PD000918; Link; 4.
 DR DR SMART: SM00032; CCP; 1.
 DR DR SMART: SM00034; CLECT; 1.
 DR DR SMART: SM00181; EGF; 1.
 DR DR SMART: SM00409; IG; 1.
 DR DR SMART: SM00445; Link; 4.
 DR DR PROSITE: PS00022; EGF_1; 1.
 DR DR PROSITE: PS01186; EGF_2; 1.
 DR DR PROSITE: PS00290; IG_MHC; 1.
 DR DR PROSITE: PS01241; LINK; 3.
 DR DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 DR DR PROSITE: PS00041; C-TYPE_LECTIN_2; 1.
 DR KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 DR KW EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
 DR FT SIGNAL 1 19 POTENTIAL.
 DR FT CHAIN 20 2415 AGGREGAN CORE PROTEIN.
 DR FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
 DR FT DOMAIN 170 247 LINK 1.
 DR FT DOMAIN 268 349 LINK 2.
 DR FT DOMAIN 495 572 LINK 3.
 DR FT DOMAIN 593 673 LINK 4.
 DR FT DOMAIN 2164 2199 EGF-LIKE.
 DR FT DOMAIN 2201 2327 C-TYPE LECTIN.
 DR FT DOMAIN 2331 2389 SUSHI.
 DR FT DOMAIN 48 141 GL-A.
 DR FT DOMAIN 152 247 GL-B.
 DR FT DOMAIN 253 349 GL-B'.
 DR FT DOMAIN 477 571 GL-B.
 DR FT DOMAIN 578 672 GL-B'.
 DR FT DOMAIN 676 848 KS.
 DR FT DOMAIN 772 844 12 X APPROXIMATE TANDEM REPEATS.
 DR FT DOMAIN 851 1497 CS-1.
 DR FT DOMAIN 941 1497 29 X APPROXIMATE TANDEM REPEATS.
 DR FT DOMAIN 1498 2162 CS-2.
 DR FT DOMAIN 2163 2415 G3.
 DR FT DISULFID 51 133 BY SIMILARITY.
 DR FT DISULFID 175 246 BY SIMILARITY.
 DR FT DISULFID 199 220 BY SIMILARITY.
 DR FT DISULFID 273 348 BY SIMILARITY.
 DR FT DISULFID 297 318 BY SIMILARITY.
 DR FT DISULFID 500 571 BY SIMILARITY.
 DR FT DISULFID 524 545 BY SIMILARITY.
 DR FT DISULFID 598 672 BY SIMILARITY.
 DR FT DISULFID 621 642 BY SIMILARITY.
 DR FT DISULFID 2168 2178 BY SIMILARITY.
 DR FT DISULFID 2173 2187 BY SIMILARITY.
 DR FT DISULFID 2189 2198 BY SIMILARITY.
 DR FT DISULFID 2205 2216 BY SIMILARITY.
 DR FT DISULFID 2233 2325 BY SIMILARITY.
 DR FT DISULFID 2301 2317 BY SIMILARITY.
 DR FT DISULFID 2332 2375 BY SIMILARITY.
 DR FT DISULFID 2361 2388 BY SIMILARITY.
 DR FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 2163 2300 MISSING (IN ISOFORM 2 AND ISOFORM 3).
 FT VARSPLIC 2330 2390 MISSING (IN ISOFORM 3).
 FT CONFLICT 766 766 E -> A (IN REF. 4).
 FT CONFLICT 847 847 E -> V (IN REF. 4).
 FT CONFLICT 1928 1928 E -> A (IN REF. 2).
 FT CONFLICT 1964 1964 I -> V (IN REF. 2 AND 3).
 FT CONFLICT 2070 2070 P -> A (IN REF. 3).
 FT CONFLICT 2391 2391 A -> P (IN REF. 2 AND 3).
 SQ SEQUENCE 2415 AA; 250191 MW; 1288937E1B98C6B6 CRC64;
 Query Match 14.2%; Score 158.5; DB 1; Length 2415;
 Best Local Similarity 29.1%; Pred. No. 6.3e-06;
 Matches 52; Conservative 22; Mismatches 62; Indels 43; Gaps 11;
 QY 9 QPVCRGG---TORPCYKVIYFHDTSRRRLNFEEAKACRRDGGQLVSISEDEOKLIEKFI 65
 DB 2202 QEVCEEGWKNKYQHCYR--HFPD---RETWVDAERRCREQSHLSIVTPEEQ----EFV 2252
 QY 66 ENLLPSDGDGF-WIGLRRREKQSNSTACQDLYAWTDGSIISQFNMYVDEPS---CGSEVC 121
 DB 2253 NN---NAQDYQWIGL-----NDRTIEGDFRWDGHPQFNRPQDNFFAAGEDC 2301
 QY 122 VYM-YHPSAPAGIGGYPYFQWDDRCNKNFNICKYS----DEKPAVPSREAEGETE 175
 DB 2302 VVMIIWHEG-----ENWDVPCNVHLPTCKTKTACGCEPPVVEHARTFGQKDK 2349
 RESULT 14
 PGCA_CANFA STANDARD; PRT; 2333 AA.
 ID PGCA_CANFA 028310; AC 028343; 028310; 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 DE AGC1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Glant T.T., Adams M.E., Kwok S.X.F., Huang D., Fulop C.;
 RT "Complete coding sequence and deduced amino acid sequence of aggrecan of canine cartilage."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 774-833 FROM N.A.
 RC TISSUE-Cartilage;
 RX MEDLINE=9512852; PubMed=7827755;
 RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
 RT "Length variation in the keratan sulfate domain of mammalian aggrecan."
 RL Matrix Biol. 14:323-328(1994).
 RN [3]
 RP SEQUENCE OF 1830-2333 FROM N.A.
 RA Adams M.E., Kwok S.X.F., Huang D., Glant T.T., Fulop C.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 2083-2118 FROM N.A.
 RC TISSUE-Cartilage;
 RX MEDLINE=9332525; PubMed=8349621;
 RA Fueleop C., Walcz E., Valyon M., Glant T.T.;
 RT "Expression of alternatively spliced epidermal growth factor-like domains in aggrecans of different species. Evidence for a novel module."
 RL J. Biol. Chem. 268:17377-17383(1993).
 CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN

RA Becker-Andre M., van Huijsduijn R.H., Losberger C., Whelan J.,
RA Delamarter J.F.;
RT "Murine endothelial leukocyte-adhesion molecule 1 is a close
RT structural and functional homologue of the human protein.";
RL Eur. J. Biochem. 206:401-411(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92340571; PubMed=1378846;
RA Weller A., Isenmann S., Vestweber D.;
RT "Cloning of the mouse endothelial selectins. Expression of both B-
RT and P-selectin is inducible by tumor necrosis factor alpha.";
RL J. Biol. Chem. 267:15176-15183(1992).
CC -/- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOLIPIDS).
CC -/- SUBCELLULAR LOCATION: Type I membrane protein.
CC -/- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.
CC -/- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -/- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -/- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80778; AAA37547.1; -;
DR EMBL; M87862; AAA37577.1; ALT_INIT.
DR HSSP; P16581; 1KJA.
DR MGD; MGI:98278; Sele.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001304; Lectin.C.
DR InterPro: IPR002396; Selectin.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00059; lectin_c; 1.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 6.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00411; C_TYPE_LLECTIN_2; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sushi; Repeat.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 612 E-SELECTIN.
FT DOMAIN 22 557 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 558 579 POTENTIAL.
FT DOMAIN 580 612 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 139 175 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 179 239 EGF-LIKE.
FT DOMAIN 242 301 SUSHI 1.
FT DOMAIN 304 364 SUSHI 2.
FT DOMAIN 367 427 SUSHI 3.
FT DOMAIN 430 490 SUSHI 4.
FT DOMAIN 493 549 SUSHI 5.
FT DOMAIN 549 612 SUSHI 6.
FT DISULFID 40 138 BY SIMILARITY.
FT DISULFID 111 130 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 180 225 BY SIMILARITY.
FT DISULFID 210 238 BY SIMILARITY.
FT DISULFID 243 287 BY SIMILARITY.

FT	DISULFID	273	300	BY SIMILARITY.
FT	DISULFID	305	350	BY SIMILARITY.
FT	DISULFID	336	363	BY SIMILARITY.
FT	DISULFID	368	413	BY SIMILARITY.
FT	DISULFID	399	426	BY SIMILARITY.
FT	DISULFID	431	476	BY SIMILARITY.
FT	DISULFID	462	489	BY SIMILARITY.
FT	DISULFID	494	535	BY SIMILARITY.
FT	DISULFID	521	548	BY SIMILARITY.
FT	CARBOHYD	25	25	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	192	192	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	203	203	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	266	266	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	391	391	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	504	504	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	612 AA;	66749 MW;	86F05713F0EC2C3D CRC64;

Query Match 13.9%; Score 154.5; DB 1; Length 612;
Best Local Similarity 27.9%; Pred. No. 2.8e-06;
Matches 38; Conservative 28; Mismatches 45; Indels 25; Gaps 5;

QY	25	YFDTSRRLNFEAKEACRRDGGOLVSIESEDEQKLEKFIENLLPSDGDWFWIGLRREE	84
Db	23	YNNASSELMTYDEASVCQRDYTHLVAIQNKEE---INLNSLNKHKHSPSYWIGIRK---	76
QY	85	KQSNSTACQDLYAWTDGS---ISQFRNYYVDEPS--CGSEVVCVVMYHQPAPAGIGGPPYM	139
Db	77	-----VNVVWIVWVGTKPLTEAQNWPGEPNKQRNEDCVETIYIQTQKDSGM-----	124
QY	140	FQWDDRCNNKNNFIC	155
Db	125	--WDERCNKKKLALC	138

Search completed: May 20, 2003, 17:05:30
Job time : 19 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:02:46 ; Search time 31 Seconds
(without alignments)
1369.216 Million cell updates/sec

Title: US-09-887-855-5

Perfect score: 1115
Sequence: 1 ATGRLSGQVCRGGTQRPC.....EDAKTKPKESREALNLAY 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1115	100.0	374	4	Q96nf3 homo sapien
2	1115	100.0	374	4	Q8tay8 homo sapien
3	1111	99.6	374	4	Q96nc5 homo sapien
4	945.5	84.8	374	11	Q92209 cricetus
5	574	51.5	273	11	Q8vi31 mus musculus
6	185	16.6	1290	13	Q9w6e1 gallus gall
7	178.5	16.0	1456	11	Q61830 mus musculus
8	177.5	15.9	1348	5	Q25199 hydra atten
9	176.5	15.8	1479	4	Q9y5p9 homo sapien
10	176.5	15.8	1479	4	Q9ubg0 homo sapien
11	176	15.8	315	4	Q9upk6 homo sapien
12	176	15.8	1321	4	O14594 homo sapien
13	170.5	15.3	1479	11	Q64449 mus musculus
14	161.5	14.5	134	5	Q9xyx3 hydra magni
15	159.5	14.3	742	11	Q8vif6 mus musculus
16	158.5	14.2	719	6	O62623 bos taurus

17	158	14.2	295	11	Q912w4 mus musculus
18	158	14.2	311	11	Q9d8v4 mus musculus
19	158	14.2	325	11	Q912x0 mus musculus
20	157	14.1	158	13	Q90w17 bungarus fa
21	156	14.0	339	6	Q95244 sus scrofa
22	152.5	13.9	1152	13	Q90wm2 xenopus lae
23	152.5	13.7	485	6	Q951g3 odocoileus
24	152	13.6	158	13	Q90w16 bungarus mu
25	150	13.5	197	6	Q28008 bos taurus
26	149.5	13.4	459	5	Q22136 caenorhabdi
27	149	13.4	158	13	Q90w18 bungarus fa
28	148.5	13.3	195	5	Q27340 megabalanus
29	148.5	13.3	330	4	Q9nt67 homo sapien
30	148.5	13.3	911	4	Q9hbk4 homo sapien
31	148.5	13.3	911	4	Q96gw7 homo sapien
32	148.5	13.3	911	4	Q96fp7 homo sapien
33	148.5	13.3	911	4	Q8tbb9 homo sapien
34	148	13.3	253	5	Q20665 caenorhabdi
35	148	13.3	381	6	Q8sqb2 macaca mula
36	147	13.2	196	11	Q9epw4 mus musculus
37	147	13.2	381	6	Q951c6 macaca neme
38	147	13.2	404	6	Q95j96 macaca mula
39	146.5	13.1	152	13	Q9dgc39 agkistrodon
40	146.5	13.1	162	5	Q25459 megabalanus
41	146.5	13.1	1031	5	Q8wsx2 dugesia tig
42	146	13.1	2109	13	P79787 gallus gall
43	145.5	13.0	134	5	Q9ub05 hydra atten
44	145.5	13.0	381	6	Q951a8 macaca mula
45	145.5	13.0	404	6	Q95198 pan troglod

ALIGNMENTS

RESULT 1

Q96nf3 PRELIMINARY; PRT; 374 AA.

AC Q96nf3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CDNA FLJ30977 fis, clone HHDPC200095, highly similar to Cricetulus griseus layilin mRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y., Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.; "NEDO human cDNA sequencing project.";
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055539; BAB70946.1;
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin-c; 1.
DR PROSITE; PS0041; C_TYPE_LECTIN_2; 1.
SQ SEQUENCE 374 AA; 42280 MW; 8A564E6BC9E56DCD CRC64;

Query Match 100.0%; Score 1115; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.8e-98;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLSGQVCRGGTQRPCYKVIYFHDTSRLNFEAEACRRDGGQLVSISEDEQKL 60
DB 22 ATGRLSGQVCRGGTQRPCYKVIYFHDTSRLNFEAEACRRDGGQLVSISEDEQKL 81

Qy 61 IEKFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 120
Db 82 IEKFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 141
Qy 121 CVVMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 180
Db 142 CVVMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 201
Qy 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227

RESULT 2
Q8TAY8 PRELIMINARY; PRT; 374 AA.
AC Q8TAY8;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Similar to unnamed protein product.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC025407; AAL25407.1; -.
SQ SEQUENCE 374 AA; 42312 MW; FC214E6BC9E578D9 CRC64;

Query Match 100.0%; Score 1115; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.8e-98;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGRLLSGPVCRCGTQPCYKVIYFHDTSRLNFEAKACRRDGGQGLVSISEDEQKL 60
Db 22 ATGRLLSGPVCRCGTQPCYKVIYFHDTSRLNFEAKACRRDGGQGLVSISEDEQKL 81
Qy 61 IEKFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 120
Db 82 IEKFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 141
Qy 121 CVVMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 180
Db 142 CVVMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 201
Qy 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227

RESULT 3
Q96NC5 PRELIMINARY; PRT; 374 AA.
AC Q96NC5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CDNA FLJ31092 fis, clone IMR321000158, highly similar to Cricetus
DE griseus layilin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saiko K., Nishikawa T., Kimura K., Yamashita H.,

Qy 61 IEKFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 120
Db 82 IEKFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 141
Qy 121 CVVMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 180
Db 142 CVVMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 201
Qy 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227

RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanchori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK055654; BAB70978.1; -.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; lectin_c; 1.
DR PROSITE: PS50041; C_Type_Lectin_2; 1.
SQ SEQUENCE 374 AA; 42310 MW; CBF74E676E23BA16 CRC64;

Query Match 99.6%; Score 1111; DB 4; Length 374;
Best Local Similarity 99.5%; Pred. No. 9.2e-98;
Matches 205; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGRLLSGPVCRCGTQPCYKVIYFHDTSRLNFEAKACRRDGGQGLVSISEDEQKL 60
Db 22 ATGRLLSGPVCRCGTQPCYKVIYFHDTSRLNFEAKACRRDGGQGLVSISEDEQKL 81
Qy 61 IEKFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 120
Db 82 IEKFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 141
Qy 121 CVVMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 180
Db 142 CVVMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 201
Qy 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227

RESULT 4
Q92209 PRELIMINARY; PRT; 374 AA.
AC Q92209;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Layilin.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Borowsky M.L., Hynes R.O.;
RT "Layilin, a novel talin-binding transmembrane protein homologous with
RT C-type lectins, is localized in membrane ruffles.";
RL J. Cell Biol. 143:0-0(1998).

DR EMBL: AF093673; AAC68695.1; -.
DR HSP; P06734; IHLI.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS50041; C_Type_Lectin_2; 1.
SQ SEQUENCE 374 AA; 42435 MW; 298A8BA24FB04E1C CRC64;

Query Match 84.8%; Score 945.5; DB 11; Length 374;
Best Local Similarity 85.9%; Pred. No. 5.5e-82;
Matches 176; Conservative 9; Mismatches 19; Indels 1; Gaps 1;

Qy 3 GRLLSGPVCRCGTQPCYKVIYFHDTSRLNFEAKACRRDGGQGLVSISEDEQKLIE 62
Db 24 GRLLSGPVCRCGTQPCYKVIYFHDTSRLNFEAKACRRDGGQGLVSISEDEQKLIE 83
Qy 63 KFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 122
Db 84 KFTIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 143
Qy 123 VMYHQPAPAGIGGPFYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPV 181

```

Db 144 VMYHOPSAPPGIGGYMFOWNDRCKMKNFICKYADEKPSITPSIRPGGEATEPTPV 203
QY 182 PEETOEDAKKTFKESREALNAY 206
Db 204 PEETOKEDKTFKESREALNAY 228

RESULT 5
OBV131
ID Q8V131 PRELIMINARY; PRT; 273 AA.
AC Q8V131;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE C-type lectin protein WT75.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Weng L., Smits P., Hubner R., Wouters J., Merregaert J.;
RT "Mt75, a low expressed c-type lectin gene involving in
chondrogenesis";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF311699; AAL50354.1;
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
KW Lectin.
SQ SEQUENCE 273 AA; 30303 MW; E052D933F244F4C7 CRC64;

Query Match 51.5%; Score 574; DB 11; Length 273;
Best Local Similarity 57.4%; Pred. No. 9.9e-47;
Matches 109; Conservative 26; Mismatches 41; Indels 14; Gaps 5;

QY 4 RLLSQPVCRGCTQRPCKYVIFHDTSRRLNFEFEAKACRRDGGQLVSTSEDEQKLIK 63
Db 23 RVVSQKVCADVAKHPCKYKMAVFEHLSRVSFQEARLACESEGGVLLSENAEOKLIES 82
QY 64 FTENLLP-----SDGFWIGLRRREKOSNSTACODLYAWTDGSIQFRNYYVDPS 118
Db 83 MLQNLTKPTGTSIDGDFWIGLRLSGDGT-SGACPDLYQWSDGSSSQFRNYYTDPSCS 141
QY 119 EVCVYVYHOPSAPGIGGYMFOWNDRCKMKNFICKYSDE-KPAPVPSREAGETELT 177
Db 142 EKCVMYVHOPTANPLGGLGYLYQWNDRCNMKNHYICTYPEIHPTPEA-----EK 196
QY 178 TPVLPETQE 187
Db 197 NQ--PEETHE 204

RESULT 6
Q9W6E1
ID Q9W6E1 PRELIMINARY; PRT; 1290 AA.
AC Q9W6E1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Neurocan core protein.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Li H., Leung T.C., Hoffman S., Balsamo J., Lillen J.;

```

```

RT *Coordinate Regulation of Cadherin and Integrin Function by the
Chondroitin Sulfate Proteoglycan Neurocan.*;
RL J. Cell Biol. 149:1275-1288(2000).
DR EMBL: AF118856; AAD24546.2;
DR HSSP: P08709; IBF9.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001438; EGF-II.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000538; Link.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00059; lectin_c; 1.
DR Pfam: PF00084; sushi; 1.
DR Pfam: PF00193; Xlink; 2.
DR PRINTS: PR00010; EGFBL00D.
DR ProDom: PD000918; Link; 2.
DR SMART: SM00032; CCP; 1.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_Like; 1.
DR SMART: SM00409; IG_1.
DR SMART: SM00445; LINK; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS00041; C_TYPE_LLECTIN_2; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS01241; LINK; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 1290 AA; 138877 MW; 182BD86D0E40BE78 CRC64;

Query Match 16.6%; Score 185; DB 13; Length 1290;
Best Local Similarity 34.7%; Pred. No. 7.7e-09;
Matches 50; Conservative 18; Mismatches 42; Indels 34; Gaps 8;

QY 17 QRPCKYVIFHDTSRRLNFEFEAKACRRDGGQLVSTSEDEQKLIKFIENLLPSDGD 76
Db 1064 QGHCYR--YF---SRRSWEDAERDCRRAGHLTSHSQEHGFINSF-----GH 1112
QY 77 IGLRREEKQSNSTACODLYAWTDGSIQFRNYYVDPS---CGSEVCVYVY-HQPSA 132
Db 1113 IGLNDRIVEQD-----FQWTDNTGLQYENWRENQPNFFAGGDCVVLVSHE 1159
QY 133 GIGGYPYMFOWNDRCKMKNFICK 156
Db 1160 -IG-----RWNDVPCYNLPYICK 1177

RESULT 7
Q61830
ID Q61830 PRELIMINARY; PRT; 1456 AA.
AC Q61830;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Macrophage mannose receptor precursor.
GN MRC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=93043353; PubMed=1421407;
RA Harris N., Rits M., Chang G., Ezekowitz R.B.;

```



```
SQ SEQUENCE 1479 AA; 166669 MW; 9F4BAF355F036FCE CRC64;
Query Match 15.8%; Score 176.5; DB 4; Length 1479;
Best Local Similarity 30.4%; Pred. No. 5.8e-08;
Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;

QY 7 SGQPVCRGGTQPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSISEDEQKLEKIE 66
Db 385 SWQPF-----QGHCYRL-----QAEKRSWQESKKACLRGGDLVSIHMAELEFITK 434
QY 67 NLLPSDGDWFGLRRREKOSNSTACQDLYAWTDGSIISQFRNWTYDEPS---CGSEVCV 123
Db 435 QEVE---ELWIGL-----NDLKLQMNFEWSDGSLVSFTWHHPFPNFRDSDCVT 483
QY 124 MYHOPAPAGIGGYMFQWDDRCNMKNFNICKYDEKPAVPSPREAG 171
Db 484 IW-----GPEG-----RWNDSPCNQSLPSICKKAGOLSOGAAEDHG 520

RESULT 10
Q9UBG0
ID Q9UBG0 PRELIMINARY; PRT; 1479 AA.
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Urokinase receptor-associated protein UPARAP.
GN KIAA0709.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA Behrendt N., Jensen O.N., Engelholm L.H., Mortz E., Mann M., Dano K.;
RP "A urokinase receptor-associated protein with specific collagen-
RT binding properties.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL: AF107292; RAF14192.1; -
DR EMBL: AB014609; BAA31684.1; -
DR HSSP: P02751; 2FN2.
DR InterPro: IPR000562; FN_Type_II.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000566; Lipocln_cytFABP.
DR InterPro: IPR000772; Ricin_B_lectin.
DR Pfam: PF00040; fn2; 1.
DR Pfam: PF00059; lectin_c; 8.
DR PRINTS: PR00013; FNTYPEII.
DR ProDom: PD000995; FN_Type_II; 1.
DR SMART: SM00034; CLECT; 8.
DR SMART: SM00059; FN2; 1.
DR SMART: SM00458; RICIN; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 3.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 8.
DR PROSITE: PS00023; FIBRONECTIN_2; UNKNOWN_1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE: PS50231; RICIN_B_LLECTIN; 1.
KW Kinase; Receptor.
SQ SEQUENCE 1479 AA; 166654 MW; C7583EA78E2792D1 CRC64;
Query Match 15.8%; Score 176.5; DB 4; Length 1479;
Best Local Similarity 30.4%; Pred. No. 5.8e-08;
Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;

SQ SEQUENCE 1479 AA; 166669 MW; 9F4BAF355F036FCE CRC64;
Query Match 15.8%; Score 176.5; DB 4; Length 1479;
Best Local Similarity 30.4%; Pred. No. 5.8e-08;
Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;

QY 7 SGQPVCRGGTQPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSISEDEQKLEKIE 66
Db 385 SWQPF-----QGHCYRL-----QAEKRSWQESKKACLRGGDLVSIHMAELEFITK 434
QY 67 NLLPSDGDWFGLRRREKOSNSTACQDLYAWTDGSIISQFRNWTYDEPS---CGSEVCV 123
Db 435 QEVE---ELWIGL-----NDLKLQMNFEWSDGSLVSFTWHHPFPNFRDSDCVT 483
QY 124 MYHOPAPAGIGGYMFQWDDRCNMKNFNICKYDEKPAVPSPREAG 171
Db 484 IW-----GPEG-----RWNDSPCNQSLPSICKKAGOLSOGAAEDHG 520

RESULT 11
Q9UPK6
ID Q9UPK6 PRELIMINARY; PRT; 315 AA.
AC Q9UPK6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PCN_HUMAN, partial CDS (fragment).
GN NEUR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J., Avila J.,
RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Frankheim M., Amico-Reiller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomey M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of an -1 Mb region containing the MEF2B gene in
RT 19p12.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005254; AAC25581.1; -
DR HSSP: P00740; 1EDM.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00059; lectin_c; 1.
DR Pfam: PF00084; sushi; 1.
DR PRINTS: PR00010; EGFBL00D.
DR SMART: SM00032; CCP; 1.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_like; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_3.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
FT NON_TER
SQ SEQUENCE 315 AA; 36306 MW; 18132088897EDA4B CRC64;
Query Match 15.8%; Score 176; DB 4; Length 315;
Best Local Similarity 31.5%; Pred. No. 1e-08;
Matches 45; Conservative 18; Mismatches 48; Indels 32; Gaps 6;

QY 17 QRPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSISEDEQKLEKIEFNLPSDGD 76
Db 171 QAEKRSWQESKKACLRGGDLVSIHMAELEFITK 434
```


Db 90 QCHYR--YF-----AHRRWEDAEKDCRRRSGHLTSVHSPEHSFINSF-----GHENTW 138
 QY 77 IGLRREEKQNSTACQDLYAWTDGSIQFRNMYVDPS---CGSEVCVVMYHQPSPAG 133
 Db 139 IGLNDRIVERD-----FQWTDNTGLGFENWRENQPDNFFAGGDCVVMVAHESG---- 187
 QY 134 IGGPYMFQWDDRCNMKNFICK 156
 Db 188 -----RWNDVPCYNLNPYVCK 203
 RESULT 12
 O14594 PRELIMINARY; PRT; 1321 AA.
 AC O14594;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Neurocan (PGCN_HUMAN).
 GN CSFG3 OR NEUR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99013874; PubMed=9795216;
 RA Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
 RT "Characterization of the human neurocan gene, CSPG3.";
 RL Gene 221:199-205(1998).
 RN [2]
 RP SEQUENCE OF 1-990 FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,
 RA Christensen M., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
 RA Ganes J., Danganan L., Bruce R., Quan G., Montgomery M., Ow D.,
 RA Kobayashi A., Olsen A.O., Carrano A.V.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF026547; AAC80576.1; -;
 DR EMBL: AC003110; AAB86655.1; -;
 DR HSSP; P00740; IEDM.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; lectin_C; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_Ca; 1.
 DR SMART; SM00001; EGF_like; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_3.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
 SQ SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB980B8 CRC64;

Query Match 15.8%; Score 176; DB 4; Length 1321;
 Best Local Similarity 31.5%; Pred. No. 5.7e-08;

Matches 45; Conservative 18; Mismatches 48; Indels 32; Gaps 6;
 QY 17 QRCQKVIYFHDTSRLNFEAEKACRRDGGQGVLSIESDEQKLEKFIENLLPSDQDFW 76
 Db 1096 QCHYR--YF-----AHRRWEDAEKDCRRRSGHLTSVHSPEHSFINSF-----GHENTW 1144
 QY 77 IGLRREEKQNSTACQDLYAWTDGSIQFRNMYVDPS---CGSEVCVVMYHQPSPAG 133
 Db 1145 IGLNDRIVERD-----FQWTDNTGLGFENWRENQPDNFFAGGDCVVMVAHESG--- 1193
 QY 134 IGGPYMFQWDDRCNMKNFICK 156
 Db 1194 -----RWNDVPCYNLNPYVCK 1209
 RESULT 13
 Q04449 PRELIMINARY; PRT; 1479 AA.
 ID Q04449;
 AC Q64449;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Lectin lambda.
 GN MRC2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96355501; PubMed=8702911;
 RA Wu K., Yuan J., Lasky L.A.;
 RT "Characterization of a novel member of the macrophage mannose receptor
 type C lectin family.";
 RL J Biol Chem. 271:21323-21330(1996).
 DR EMBL; U56734; AAC52729.1; -;
 DR HSSP; P02751; 2FN2.
 DR MGD; MGI:107818; MRC2.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR InterPro; IPR000562; FN_Type_II.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000566; Lipoclin_cytFABP.
 DR InterPro; IPR000772; Ricin_B_lectin.
 DR Pfam; PF00040; fn2; 1.
 DR Pfam; PF00059; lectin_C; 8.
 DR PRINTS; PR00013; FNTYPEII.
 DR ProDom; PD000995; FN_Type_II; 1.
 DR SMART; SM00034; CLECT; 8.
 DR SMART; SM00059; FN2; 1.
 DR SMART; SM00458; RICIN; 1.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 3.
 DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 8.
 DR PROSITE; PS00023; FIBROECTIN_2; UNKNOWN_1.
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
 DR PROSITE; PS0231; RICIN_B_LLECTIN; 1.
 SQ SEQUENCE 1479 AA; 167112 MW; 62D456E10B9B48C1 CRC64;
 Query Match 15.3%; Score 170.5; DB 11; Length 1479;
 Best Local Similarity 31.4%; Pred. No. 2.2e-07;
 Matches 48; Conservative 22; Mismatches 48; Indels 35; Gaps 7;
 QY 7 SQQPVCRGGTQRCQKVIYFHDTSRLNFEAEKACRRDGGQGVLSIESDEQKLEKFE 66
 Db 384 SQQPF-----QGHCVRL-----QAEKRSWQESKRACLRGGGDLLSIHSMAELEFTRKIK 433
 QY 67 NLLPSDGFHWGLRRRREKQNSTACQDLYAWTDGSIQFRNMYVDPS---CGSEVCV 123
 Db 434 QEVE---ELWIGL-----NDLKLQMFNWSQGLSVSFTHWHFEPNPFDSLEDCVT 482
 QY 124 MYHQPSPAGIGGPMFQWDDRCNMKNFICK 156
 Db 483 IW-----GPEG-----RWNDSPCNQSLPSICK 504

```

RESULT 14
Q9XYX3 PRELIMINARY; PRT; 134 AA.
AC Q9XYX3
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor protein-tyrosine kinase (Fragment).
GN HTK28.
OS Hydra magnipapillata (Hydra).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6085;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=105;
RX MEDLINE=20209407; PubMed=10744720;
RA Reidling J.C., Miller M.A., Steele R.E.;
RT "Sweet Tooth, a Novel Receptor Protein-tyrosine Kinase with C-type
RT Lectin-like Extracellular Domains.";
RL J. Biol. Chem. 275:10323-10330(2000).
DR EMBL; AF129528; AAD30040.1;
DR HSP; P22897; IEGG.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; Lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Kinase; Tyrosine-protein Kinase.
FT NON_TER 1
FT TER 134
SQ SEQUENCE 134 AA; 15701 MW; E7B7211C881009BC CRC64;

Query Match 14.5%; Score 161.5; DB 5; Length 134;
Best Local Similarity 28.8%; Pred. No. 8.9e-08;
Matches 40; Conservative 23; Mismatches 51; Indels 25; Gaps 5;

QY 20 CYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIIEDEQKLIKFIENLLPSDGFHWIGL 79
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 CY--FFQNKTLQAKNRDASLSCQALGGHLLSIEDQAEFFILNFKDSSWQQDNYWIGL 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 80 RRREEKQSNSTACQDLVATDGSISQFRNMYVDEP--CGSEVCVVMYHOPAPAGIGGP 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 -----NDASNNREFRWSDDKIQPFNWLPKPNNDQSQNCV-----ETNSMG-- 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 138 YMFQWDDRCNMKNFICK 156
   ||| : : : : :
Db 117 ----WDENCDATNGFICK 131

RESULT 15
Q8VIF6 PRELIMINARY; PRT; 742 AA.
AC Q8VIF6
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Scavenger receptor with C-type lectin.
GN COLEC12 OR SRCL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21575692; PubMed=11718900;
RA Nakamura K., Funakoshi H., Tokunaga F., Nakamura T.;
RT "Molecular cloning of a mouse scavenger receptor with C-type lectin
RT (SRCL)(1), a novel member of the scavenger receptor family.";
RL Blochim. Biophys. Acta 1522:53-58(2001).
DR EMBL; AB038519; BAB82497.1;

```

```

DR MGI; 2152907; Colecl2.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00059; Lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Receptor; Lectin.
SQ SEQUENCE 742 AA; 81307 MW; 85A90D3AE881DB6B CRC64;

Query Match 14.3%; Score 159.5; DB 11; Length 742;
Best Local Similarity 33.1%; Pred. No. 1.1e-06;
Matches 49; Conservative 17; Mismatches 55; Indels 27; Gaps 8;

QY 20 CYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIIEDEQKLIKFIENLLPSDGFHWIGL 79
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 618 CY---YF--SLEKEILEDAKLFCEDKSSHLVFINRSEEQWIKKH-----TVGRESHWIGL 668
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 80 RRREEKQSNSTACQDLVATDGSISQFRNMYVDEP--CGSEVCVVMYHOPAPAGIGGP 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 669 TDSEGESE-----WKWLDGSPVDYKNWKAGQPDNWSG-----HGPGECA-GLIY 713
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 139 YMFQWDDRCNMKNFICKYSDEKPAVPS 166
   ||| : : : : :
Db 714 AGQWDFQCDENNFICE--KREAVPS 739

Search completed: May 20, 2003, 17:06:04
Job time : 34 secs

```